UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Asymptotic Theory for Beta-t-GARCH

Ryoko Ito

24 January 2016

CWPE 1607

Asymptotic Theory for Beta-t-GARCH

Ryoko Ito*

Faculty of Economics, University of Cambridge

This version: January 24, 2016

ABSTRACT

The consistency and asymptotic normality of the maximum likelihood estimator is established for the first order Beta-t-GARCH model, which is a special case of the dynamic conditional score (DCS) model and closely related to the first-order Gaussian GARCH model as its limiting case. The necessary and sufficient parameter space for which the process is strictly stationary and ergodic is established, and the asymptotic results are shown to hold for the strictly stationary version of the model.

KEYWORDS: robustness; score; consistency; asymptotic normality.

JEL Classification: C22, C58

1. INTRODUCTION

This paper shows the consistency and asymptotic normality of the maximum likelihood estimator (MLE) in the first-order Beta-t-GARCH model, which is given by

$$\begin{aligned} y_t &= \gamma_0 + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_{0t}} z_t, \quad b_{0t} = \frac{z_t^2}{z_t^2 + (\nu_0 - 2)}, \\ h_{0t} &= \delta_0 + \beta_0 h_{0t-1} + \alpha_0(\nu_0 + 1) h_{0t-1} b_{0t-1}, \end{aligned} \tag{1}$$

with $t \in \mathbb{N}_{>0}$, where $(y_t)_{t\in\mathbb{N}}$ is a process generated on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. $(z_t)_{t\in\mathbb{N}_{>0}}$ is assumed to be

*Email: ri239@cam.ac.uk. I would like to thank my supervisor Andrew Harvey for his helpful guidance and comments on my postgraduate research throughout the past four years. I would also like to thank Francisco Blasques, Axel Gottfries, Oliver Linton, Andre Lucas, Vladimirs Murevics, Bent Nielsen, Anders Rahbek, and Stephen Thiele for providing thoughtful comments on preliminary versions of this paper. Finally, I would like to thank the Cambridge Trust, the International Monetary Fund, the Keynes Fund, and the Stevenson Fund (of the Faculty of Economics, Cambridge University) for providing various forms of financial support.

independently and identically distributed (i.i.d.) as the standard Student's t-distribution with ν_0 degrees of freedom and the first two moments, $\mathbb{E}[z_t] = 0$, and $\mathbb{E}[z_t^2] = 1$. $(b_{0t})_{t\in\mathbb{N}_{>0}}$ is i.i.d. and follows the beta distribution with parameters $(1/2, \nu_0/2)$ (denoted by $\text{Beta}(1/2, \nu_0/2)$) by the properties of Student's t. (The formal definition of Student's t and the beta distribution are in Appendix A.) The Beta-t-GARCH model is defined formally in Harvey (2013, p. 125).

When $\nu_0 \to \infty$, the Beta-t-GARCH model becomes the generalized autoregressive conditional heteroscedasticity (GARCH) model introduced by Bollerslev (1986) and Taylor (1986). This is because Student's t converges in distribution to the standard normal distribution and

$$(\nu_0 + 1)b_{0t} \to \varepsilon_t^2/h_{0t} = z_t^2$$

as $\nu_0 \to \infty$. This means that the conditional variance process becomes

$$h_{0t} = \delta_0 + \beta_0 h_{0t-1} + \alpha_0 \varepsilon_{t-1}^2$$

as $\nu_0 \to \infty$, giving the GARCH filter.

The quasi-maximum likelihood estimator (QMLE) is often applied to estimate GARCH. Lumsdaine (1996) and Lee and Hansen (1994) establish the consistency and asymptotic normality of QMLE when the true GARCH$(1, 1)$ process is strictly stationary. The asymptotic results are generalized to the GARCH(p, q) case by Boussama (2000), Berkes et al. (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006), also under the assumption that the true GARCH(p, q) process is strictly stationary. Jensen and Rahbek (2004) establish the consistency and asymptotic normality of QMLE when the true GARCH$(1, 1)$ process is nonstationary. Francq and Zakoïan (2010, Ch. 7 and 9) provide a comprehensive review of the theoretical development in the GARCH literature. A notable aspect of it is that the asymptotic normality of QMLE in GARCH fails (for both the strictly stationary and nonstationary cases) when the fourth moment of the error distribution is not finite. The efficiency of QMLE is also determined by how far the distribution of data is from normality. QMLE in GARCH can be inefficient and the fourth moment condition may be violated in many financial applications as financial data are often highly non-Gaussian and display a relatively high degree of kurtosis (see, for instance, Caviano and Harvey (2013a, 2013b) and Ibragimov et al. (2013)). Several authors have proposed nonparametric procedures to deal with non-normality; for instance, Hall and Yao (2003) propose the use of percentile-t subsample bootstrap to approximate the estimator's distribution when necessary conditions for asymptotic normality are violated. However, nonparametric procedures typically add computational costs to the simple estimation procedure of QMLE.

As for the MLE in GARCH, Berkes and Horváth (2004) establish that the MLE in GARCH(p, q) is consistent and asymptotically normal under certain regularity assumptions on the parameter space and the shape of the error distribution. They show asymptotic normality without assuming that the fourth moment of the error distribution is finite.

GARCH is known to lack robustness against outliers, which is undesirable if extreme observations rarely occur and should not substantially change the conclusions of a given study. Several authors have proposed robust modifications of GARCH (see, for instance, Li et al. (2010) and Park (2002)). The robustness literature typically deals with outliers by taking approaches classified as Winsorising or trimming. Many robust-GARCH models are still found to lack robustness against isolated additive outliers (see Muler and Yohai (2008)). Muler and Yohai (2008) propose the bounded M-estimation procedure to overcome the robustness issue. It places some constant upper-bound on standardized squared observations, which drive the conditional variance dynamics, in order to Winsorize the effect of extreme observations. The need to modify GARCH to ensure robustness is due to the fact that the model formulates conditional variance as a weighted linear sum of past squared observations. This is analogous to the sample variance formula, which is not an efficient estimator if the underlying distribution is highly non-Gaussian. One may naturally question whether there is an alternative class of observation-driven volatility model with a unified robustness framework that ensures that the asymptotic properties of likelihood-based estimators do not depend on the degree of non-Gaussianity of data.

Beta-t-(E)GARCH introduced by Harvey and Chakravarty (2008) is a new class of observation-driven model that takes a step in the above direction. Beta-t-(E)GARCH is a special case of the dynamic conditional score (DCS) model, formally defined and studied by Harvey (2013), and also independently by Creal et al. (2011, 2013). The latter authors call DCS the generalized autoregressive score (GAS) model. Beta-t-(E)GARCH is simple in structure, computationally practical, and robust to extreme observations because its volatility dynamics are driven by the *score*, a measure of how distribution changes as volatility evolves. In the case of Beta-t-(E)GARCH, the score has the beta distribution and Winsorizes the effect of extreme observations on the dynamics of the filter. As Beta-t-GARCH becomes the Gaussian GARCH when $\nu_0 \to \infty$, presumably Beta-t-GARCH can capture kurtosis that is in some sense "too high" for GARCH. Beta-t-GARCH is useful for testing the comparative merits of DCS and GARCH. See, for instance, Koopman et al. (2012), Harvey and Sucarrat (2012), and Ito (2013, 2016) for the practical features of DCS.

Harvey (2013) shows the consistency and asymptotic normality of MLE in the weakly stationary DCS when the scale (or volatility) parameter is modeled via the exponential link function and some useful continuous distribution is chosen for the error distribution. The asymptotic results of Harvey (2013) do not depend on high-order moment assumptions. A main objection to the ML approach is that it requires the knowledge of the error distribution. This does not seem to pose a major practical problem for DCS as the asymptotic results by Harvey (2013) extend to useful generalized distributions (such as the generalized beta distribution of the second kind) that encompass a number of well-known heavy-tailed distributions as their special cases. The asymptotic analysis by Harvey (2013) implies that MLE in Beta-t-EGARCH, a DCS model with the Student's t-distribution and the exponential link function, is consistent and asymptotically normal when conditional variance dynamics are assumed to be weakly stationary.

We establish the consistency and asymptotic normality of MLE in the first-order Beta-t-GARCH. Blasques et al. (2014) derive conditions under which the consistency and asymptotic normality of the MLE in DCS can be shown to hold. Their analysis focuses on the parameter region for which the model is strictly stationary and ergodic. We establish the necessary and sufficient condition of the parameter space for which the true Beta-t-GARCH process is strictly stationary. Then we show the consistency and asymptotic normality of MLE when the true process is strictly stationary. The asymptotic properties are derived using Lemma 1 of Jensen and Rahbek (2004).[1] As regard moment assumptions for the error distribution (i.e. Student's t), we assume only that the second moment is finite. Although this assumption is necessary as Beta-t-GARCH is a volatility model, we may be able to relax it if we reformulate

[1] A notable feature of Lemma 1 of Jensen and Rahbek (2004) is the condition (A.3), which requires the third derivative of log-likelihood to be bounded in some neighborhood of true parameter values by a process that is convergent in probability. In contrast, Lumsdaine (1996) and Lee and Hansen (1994) apply the uniform convergence results of functionals by Andrews (1987 or 1992), and the convergence results, described in Amemiya (1985), for the maximizer of a function defined over a compact parameter space. The use of the results in Amemiya (1985) in establishing asymptotic normality require the authors to prove that the third-derivative of log-likelihood is bounded in L^1 over some admissible parameter region.

it as a model for time-varying scale instead of volatility. Our asymptotic results imply that the MLE in Beta-t-GARCH can be applied to data with a relatively high degree of kurtosis or nonstationary features without compromising the asymptotics of MLE.

Our analysis imposes very mild assumptions on the parameter space: the unconditional mean of observations ($\mathbb{E}[y_t] \equiv \gamma_0$) is finite, all of the parameters of the Beta-t-GARCH filter are bounded and strictly positive (so that the volatility dynamics remain positive), and $2 < \nu_0 < \infty$ so that the second moment of Student's t is finite and the Gaussian GARCH case is excluded from our inference. The unit upper-bound on the dynamic parameter, β_0, which is the coefficient on the lagged conditional variance, is often assumed in the asymptotic inference for either the QMLE or MLE in GARCH when the true process is assumed to be strictly stationary (see, for instance, Lee and Hansen (1994) and Berkes et al. (2003)).[2] We consider the parameter space for which $\beta_0 < 1$. We use the unit upper-bound to show that the third derivative of the log-likelihood function is bounded by some strictly stationary process (see Lemmas 15-16 in Appendix D). In the strictly stationary case, apart from the initial volatility parameter (denoted by $\omega_0 \equiv h_{00}$ later on), consistency and asymptotic normality are established for all parameters of Beta-t-GARCH including the unconditional mean of observations (γ_0) and the intercept (δ_0) of the filter. We find that the parameters for the unconditional mean of observations (γ_0) and the degrees of freedom (ν_0) are the most difficult to handle. In the nonstationary case, we conjecture that the asymptotic results can be established for α_0, β_0, and ν_0. We think that the asymptotic results for the non-stationary case would not hold for (δ, γ) in the nonstationary case since the second derivative of the log-likelihood with respect to δ or γ collapses to zero asymptotically (see Lemma 11).[3]

[2]The parameter space of Lee and Hansen (1994) impose a unit upper-bound on β_0. The consistency of global QMLE also requires $\alpha_0 + \beta_0 < 1$, where α_0 and β_0 are the coefficients on lagged squared observation and the lagged conditional variance, respectively. The (strong) consistency of QMLE in GARCH(p, q) require that the GARCH coefficients on the lags of conditional variance sum to less than one (i.e. $\sum_{j=1}^{p} \beta_{0j} < 1$), which is implied by the strict stationarity assumption (Corollary 2.3 of Bougerol and Picard (1992)).

[3]This compares with the results by Jensen and Rahbek (2004), which require γ_0, δ_0 and ω_0 to be fixed and known. The QMLE of δ_0 is found to be inconsistent in nonstationary GARCH$(1, 1)$. (See the discussions in Jensen and Rahbek (2004) and Francq and Zakoïan (2010, p. 180)).

2. STATIONARITY AND ERGODICITY

Our analysis is conditional on the initial value $y_0 \in \mathbb{R}$. The initial value of the conditional variance, h_{0t}, is parameterized by $h_{00} = \omega_0 \in \mathbb{R}_{>0}$. The vector of true parameters are denoted by $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top \in \Theta$, where

$$\Theta = \{\theta \in \mathbb{R}^6 : 2 < \nu_u \leq \nu \leq \nu_u < \infty, 0 < \alpha_l \leq \alpha \leq \alpha_u < \infty, 0 < \beta_l \leq \beta \leq \beta_u < 1,$$
$$0 < \delta_l \leq \delta \leq \delta_u < \infty, -\infty < \gamma_l \leq \gamma \leq \gamma_u < \infty, 0 < \omega_l \leq \omega \leq \omega_u < \infty\}.$$

Assuming that θ_0 is unknown, we estimate the model,

$$\begin{aligned} y_t &= \gamma + e_t, \quad h_t(\theta) = \delta + \beta h_{t-1}(\theta) + \alpha(\nu+1)h_{t-1}(\theta)b_{t-1}(\theta), \\ b_t(\theta) &= \frac{e_t^2}{e_t^2 + (\nu-2)h_t(\theta)}, \end{aligned} \tag{2}$$

for $\theta \in \Theta$, where $t \in \mathbb{N}_{\geq 0}$ and $\theta = (\nu, \alpha, \beta, \delta, \gamma, \omega)^\top \in \Theta$. At $\theta = \theta_0$, we have $h_t(\theta_0) = h_{0t}$ and $b_t(\theta_0) = b_{0t}$ so that $(h_t(\theta))_{t\in\mathbb{N}_{>0}}$ follows the dynamics specified in (1) at $\theta = \theta_0$.

When $\nu_0 \to \infty$, Nelson (1990) shows that $\mathbb{E}[\ln(\beta_0 + \alpha_0 z_t^2)] < 0$ is the necessary and sufficient condition for the true GARCH$(1,1)$ process (i.e. Beta-t-GARCH with large ν_0) to be strictly stationarity. To establish the strict stationarity condition for the true Beta-t-GARCH$(1,1)$ process with any $\nu_0 \in (2, \infty)$, we split Θ into two regions;

$$\Theta_L \equiv \{\theta_0 \in \Theta : \mathbb{E}\left[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})\right] < 0\},$$
$$\Theta_U \equiv \{\theta_0 \in \Theta : \mathbb{E}\left[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})\right] \geq 0\}.$$

We denote the convergence in probability by $\xrightarrow{P}$ and in distribution by $\xrightarrow{\mathcal{D}}$. We use $\|\cdot\|_p$ for $p \geq 1$ to denote the L^p-norm on $(\Omega, \mathcal{F}, \mathbb{P})$.

THEOREM 1. *If $\theta_0 \in \Theta_L$, $(h_{0t})_{t\in\mathbb{N}_{>0}}$ is strictly stationary and ergodic with a well-defined probability measure μ_∞ on (δ_0, ∞). If $\theta_0 \in \Theta_U$, $(h_{0t})_{t\in\mathbb{N}_{>0}}$ is divergent almost surely (a.s.) and its reciprocal converges to zero a.s. as well as in L^p for any $p \geq 1$ as $t \to \infty$.*

In what follows, the ith element of $\theta = (\nu, \alpha, \beta, \delta, \gamma, \omega)^\top \in \Theta$ may be denoted by θ_i for $i = 1, 2, \ldots, 6$, so that $\theta_1 \equiv \nu$ and so on. Define

$$h_{\theta_i t}(\theta) \equiv \frac{\partial h_t(\theta)}{\partial \theta_i}\frac{1}{h_t(\theta)}, \quad h_{\theta_i\theta_j t}(\theta) \equiv \frac{\partial^2 h_t(\theta)}{\partial\theta_i\partial\theta_j}\frac{1}{h_t(\theta)}, \quad h_{\theta_i\theta_j\theta_k t}(\theta) \equiv \frac{\partial^3 h_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\frac{1}{h_t(\theta)}$$

for $i, j, k = 1, 2, \ldots, 6$, so that $h_{\beta t}(\theta) = h_t(\theta)^{-1}(\partial h_t(\theta)/\partial\beta)$ and so on. The analytic expressions for the derivatives of the log-likelihood, as well as $h_{\theta_i t}(\theta)$, $h_{\theta_i\theta_j t}(\theta)$, and $h_{\theta_i\theta_j\theta_k t}(\theta)$ for $i, j, k = 1, 2, \ldots, 6$ are given in Appendix B.1. We also set $\prod_{j=1}^{0} \cdot = 1$ for notational convenience.

Next, we consider the stationarity property of the log-likelihood function. Given a finite sequence of observations $(y_t)_{t=1}^{n}$ for some $n \in \mathbb{N}_{>0}$, the log-likelihood for the Beta-t-GARCH model is

$$L_n(\theta) = n^{-1} \sum_{t=1}^{n} l_t(\theta),$$

where

$$\begin{aligned} l_t(\theta) \equiv \ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}\ln(\nu-2) - \frac{1}{2}\ln(\pi) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) \\ - \frac{1}{2}\ln\left(h_t(\theta)\right) - \frac{\nu+1}{2}\ln\left(1 + \frac{e_t^2}{(\nu-2)h_t(\theta)}\right). \end{aligned}$$

Theorem 2 establishes the stationarity and ergodicity properties of the log-likelihood function and its first two derivatives with respect to θ evaluated at $\theta = \theta_0 \in \Theta_L$.

THEOREM 2. *If $\theta_0 \in \Theta_L$, $(l_t(\theta_0))_{t\in\mathbb{N}}$ and its first two derivatives of $(l_t(\theta))_{t\in\mathbb{N}}$ with respect to θ evaluated at $\theta = \theta_0$, denoted by $(\nabla_\theta l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_\theta^2 l_t(\theta_0))_{t\in\mathbb{N}}$, are strictly stationary and ergodic.*

3. CONSISTENCY AND ASYMPTOTIC NORMALITY OF MLE

We assume that the true initial value of the volatility process is known (i.e. $h_0(\theta) \equiv \omega = \omega_0 \equiv h_{0t}$) throughout Section 3. Thus, throughout Section 3, we reduce the dimension of the parameter space to

$$\theta = (\nu, \alpha, \beta, \delta, \gamma)^\top \in \Theta \subset \mathbb{R}^5,$$

where the dimension of Θ is adjusted accordingly. We write $L_n(\theta) \equiv L_n(\delta, \alpha, \beta, \gamma, \nu, \omega_0)$, and likewise for the single log-likelihood function, $l_t(\cdot)$, and $h_t(\cdot)$.

Theorem 3 states that, when $\theta_0 \in \Theta_L$ and $\omega = \omega_0$, the MLE of θ_0 is consistent and asymptotically normal.

Define $Q(\theta_0) \equiv \mathbb{E}[\nabla_\theta^2 l_t(\theta_0)]$ and $R(\theta_0) \equiv \mathbb{E}[\nabla_\theta l_t(\theta_0) \nabla_\theta l_t(\theta_0)^\top]$. The existence of these moments are by Lemmas 6 and 9. These notations mean that the derivatives of the log-likelihood function are taken with respect to the free parameters only, i.e.

$$\nabla_\theta l_t(\theta_0) = (\partial l_t(\theta_0)/\partial\nu, \partial l_t(\theta_0)/\partial\alpha, \partial l_t(\theta_0)/\partial\beta, \partial l_t(\theta_0)/\partial\delta, \partial l_t(\theta_0)/\partial\gamma).$$

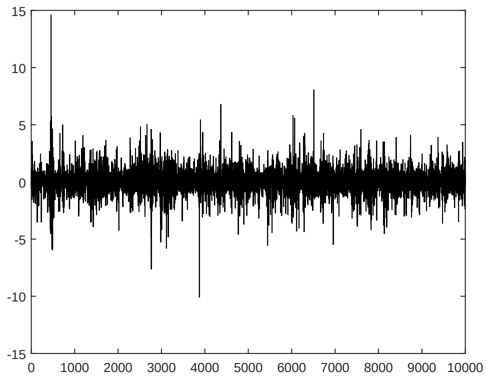

Figure 1 The time series plot of y_t when $n = 10,000$ and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.13, 0.84, 0.03, 0.05, 1)^\top$. Time on the x-axis.

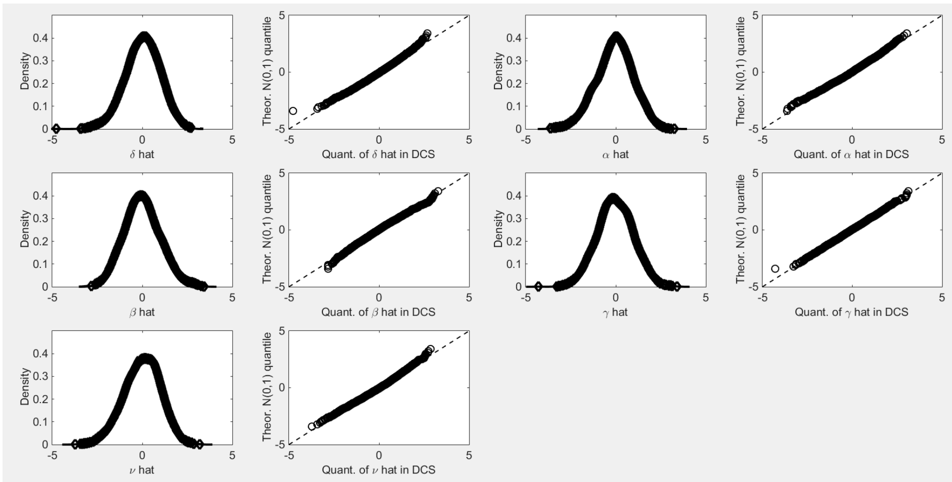

Figure 2 The empirical distribution (the first and the third columns) and the QQ-plot (the second and the fourth columns) of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$, which is standardized by $n^{-1}\sum_{t=1}^{n}(\partial L_n(\theta_0)/\partial\theta)(\partial L_n(\theta_0)/\partial\theta^\top)$. $n = 10,000$, $K = 3,000$, and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.13, 0.84, 0.03, 0.05, 1)^\top$. From the top left panel, $\widehat{\delta}$, $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\gamma}$, and $\widehat{\nu}$.

THEOREM 3. *Suppose θ_0 is an interior point of Θ. Assume that $\theta_0 \in \Theta_L$. Then, with probability tending to one, there exists a unique maximum point $\widehat{\theta}_n$ of $L_n(\theta)$ such that $\widehat{\theta}_n \xrightarrow{\mathcal{P}} \theta_0$ and*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, V(\theta_0))$$

as $n \to \infty$, where $V(\theta_0) \equiv Q(\theta_0)^{-1}R(\theta_0)Q(\theta_0)^{-1} = Q(\theta_0)^{-1}$.

In this paper, we do not show that the above results are asymptotically independent of the value of ω in the strictly stationary case. We will consider this in our future research.

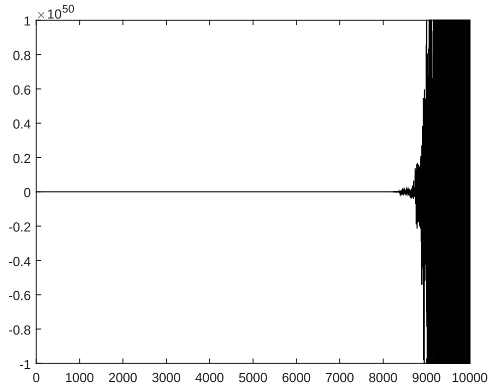

Figure 3 The time series plot of y_t when $n = 10,000$ and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.16, 0.88, 0.04, 0.05, 1)^\top$. Time on the x-axis.

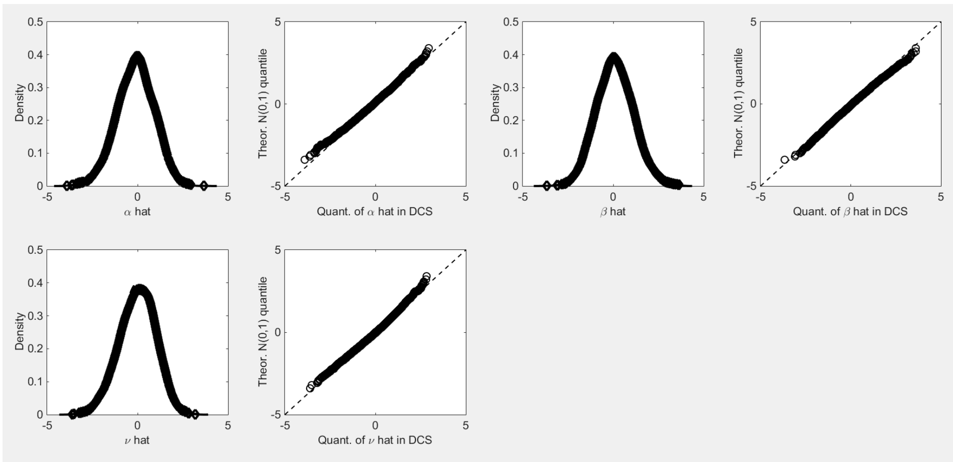

Figure 4 The empirical distribution (the first and the third columns) and the QQ-plot (the second and the fourth columns) of $\sqrt{n}(\widehat{\theta}^*_n - \theta^*_0)$, which is standardized by $n^{-1}\sum_{t=1}^{n}(\partial L_n(\theta_0)/\partial\theta^*)(\partial L_n(\theta_0)/\partial\theta^{*\top})$. $n = 10,000$, $K = 3,000$, and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.16, 0.88, 0.04, 0.05, 1)^\top$. From the top left panel, $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\nu}$.

4. Simulation results

We simulate the asymptotic distribution of $\widehat{\theta}_n$ and $\widehat{\theta}^*_n$ for when $\theta_0 \in \Theta_L$ and $\theta_0 \in \Theta_U$, respectively. Figure 1 shows the time series of $(y_t)_{t=1}^n$ when $n = 10,000$ and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0) = (5, 0.13, 0.84, 0.03, 0.05, 1)$.[4] The simulation size is $K = 3,000$. Although we did not have the analytical moment expression for $\mathbb{E}[\log(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t})]$, we should have $\theta_0 \in \Theta_L$. For the first simulation, the sample mean of $(\log(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}))_{t=1}^T$ was -0.0413, which is statistically

[4]The selected values of θ_0 are close to typical parameter estimates in empirical application. For instance, if we fit the first-order Beta-t-GARCH model to the daily returns (computed using the daily closing level) of the Dow Jones Industrial Average index between 1 October 1975 and 5 July 2011, we obtain $\widehat{\theta} = (\widehat{\nu}, \widehat{\alpha}, \widehat{\beta}, \widehat{\delta}, \widehat{\gamma}, \widehat{\omega})^\top = (0.04, 0.13, 0.83, 0.05, 5.09)^\top$.

significantly (and comfortably) different from zero at the 5% level given its standard deviation of 0.1390 and the sample length, $T = 10,000$. Likewise, the sample mean of $(\log(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}))_{t=1}^T$ was negative and statistically significantly (and comfortably) different from zero at the 5% level for all of the $K = 3,000$ simulations. Figure 2 shows the simulated asymptotic distribution of $\widehat{\theta}_n$ computed using these simulated time series. $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is standardized by $n^{-1}\sum_{t=1}^n(\partial L_n(\theta_0)/\partial\theta)(\partial L_n(\theta_0)/\partial\theta^\top)$ to obtain the figure. The QQ-plot compares the simulated distribution with the theoretical quantiles of the standard normal distribution. The asymptotic normality appears to hold for $\widehat{\theta}_n$.

To illustrate the asymptotic behavior of MLEs for the nonstationary case, Figure 3 shows the time series of $(y_t)_{t=1}^n$ when $n = 10,000$ and $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.16, 0.88, 0.04, 0.05, 1)^\top$. We should have $\theta_0 \in \Theta_U$. For the first simulation, the sample mean of $(\log(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}))_{t=1}^T$ was 0.0252, which is statistically significantly (and comfortably) different from zero at the 5% level given its standard deviation of 0.1582 and the sample length, $T = 10,000$. Likewise, the sample mean of $(\log(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}))_{t=1}^T$ was positive and statistically significantly (and comfortably) different from zero at the 5% level for all of the $K = 3,000$ simulations. Figure 4 shows the simulated asymptotic distribution of $\widehat{\theta}_n$ computed using these simulated time series. The estimators are standardized as before. The asymptotic normality appears to hold for the parameters shown in the figure.

REFERENCES

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

Andrews, D. W. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

— (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Berkes, I. and Horváth, L. (2004), "The Efficiency of the Estimator of the Parameters in GARCH Processes," *The Annals of Statistics*, 32, 633–655.

Berkes, I., Horváth, L., and Kokoszka, P. (2003), "Asymptotics for GARCH Squared Residual Correlations," *Econometric Theory*, 19, 515–540.

Billingsley, P. (1986), *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, Wiley: New York, 2nd ed.

Blasques, F., Koopman, S. J., and Lucas, A. (2014), "Maximum Likelihood Estimation for Generalized Autoregressive Score Models," Tinbergen Institute Discussion Paper TI2014-029/III, Tinbergen Institute.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bougerol, P. and Picard, N. (1992), "Stationarity of GARCH Processes and of Some Nonnegative Time Series," *Journal of Econometrics*, 52, 115–127.

Boussama, F. (2000), "Normalité Asymptotique de l'Estimateur du Pseudo-Maximum de Vraisemblance d'un Modéle GARCH," *Comptes Rendus de l'Académie de Sciences Paris*, 331, 81–84.

Brandt, A. (1986), "The Stochastic Equation Yn+1=AnYn+Bn with Stationary Coefficients," *Advances in Applied Probability*, 18, 211–220.

Caviano, M. and Harvey, A. C. (2013a), "Time Series Models with an EGB2 Conditional Distribution," Cambridge Working Papers in Economics CWPE1325, University of Cambridge.

— (2013b), "Two EGARCH Models and One Fat Tail," Cambridge Working Papers in Economics CWPE1326, University of Cambridge.

Creal, D. D., Koopman, S. J., and Lucas, A. (2011), "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," *Journal of Business and Economic Statistics*, 29, 552–563.

— (2013), "Generalized Autoregressive Score Models With Applications," *Journal of Applied Econometrics*, 28, 777–795.

Francq, C. and Zakoïan, J.-M. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

— (2010), *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley.

Hall, P. and Yao, Q. (2003), "Inference in ARCH and GARCH Models with Heavy-Tailed Errors," *Econometrica*, 71, 285–317.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Harvey, A. C. (2013), *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Econometric Society Monograph, Cambridge University Press.

Harvey, A. C. and Chakravarty, T. (2008), "Beta-t-(E)GARCH," Cambridge Working Papers in Economics CWPE0840, University of Cambridge.

Harvey, A. C. and Sucarrat, G. (2012), "EGARCH Models with Fat Tails, Skewness, and Leverage," Cambridge Working Papers in Economics CWPE1236, University of Cambridge.

Ibragimov, M., Ibragimov, R., and Kattuman, P. (2013), "Emerging Markets and Heavy Tails," *Journal of Banking and Finance*, 37, 2546–2559.

Ito, R. (2013), "Modeling Dynamic Diurnal Patterns in High Frequency

Financial Data," Cambridge Working Papers in Economics CWPE1315, University of Cambridge.

— (2016), "Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data," Cambridge Working Papers in Economics, University of Cambridge, forthcoming.

Jensen, S. T. and Rahbek, A. (2004), "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

Koopman, S. J., Lucas, A., and Scharth, M. (2012), "Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models," Tinbergen Institute Discussion Papers 12-020/4, Tinbergen Institute.

Lee, S.-W. and Hansen, B. E. (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Li, J., Kao, C., and Zhang, W. D. (2010), "Bounded Influence Estimator for GARCH Models: Evidence from Foreign Exchange Rates," *Applied Economics*, 42, 1437–1445.

Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575–96.

Muler, N. and Yohai, V. J. (2008), "Robust Estimates for GARCH Models," *Journal of Statistical Planning and Inference*, 138, 2918–2940.

Nelson, D. B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Park, B. (2002), "An Outlier Robust GARCH Model and Forecasting Volatility of Exchange Rate Returns," *Journal of Forecasting*, 21, 381–393.

Royden, H. L. (1988), *Real Analysis*, Macmillan: New York, 3rd ed.

Stout, W. F. (1974), *Almost Sure Convergence*, Probability and mathematical statistics, Academic Press: New York.

Straumann, D. and Mikosch, T. (2006), "Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach," *The Annals of Statistics*, 34, 2449–2495.

Taylor, S. (1986), *Modelling Financial Time Series*, Chichester West Sussex ; New York: Wiley.

Varadhan, S. R. S. (2001), *Probability Theory*, Courant Lecture Notes, Courant Institute of Mathematical Sciences.

APPENDIX A: Student's t-distribution

The standard Student's t-distribution has the probability density function (pdf),

$$f(x;\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\left(1+\frac{x^2}{(\nu-2)}\right)^{-(\nu+1)/2}, \quad x \in \mathbb{R},\ \nu > 2,\ h > 0,$$

where $\nu > 0$ is the degrees of freedom and $\Gamma(\cdot)$ is the gamma function. The mean is 0 and variance is 1.

If a random variable Y follows the standard Student's t-distribution after it is standardized by a scaling parameter $h > 0$, the pdf of Y denoted by $f_Y : \mathbb{R}_{>0} \to \mathbb{R}$ is $f_Y(y;h,\nu) = f(y/h;\nu)/h$ for $y \in \mathbb{R}$. Since the variance of the standard Student's t-distribution is normalized, the variance of Y is determined by h and not by ν. For a set of i.i.d. observations $y_1, \ldots, y_N$ where each follows the non-standardized Student's t, the log-likelihood function of a single observation y_t can be written as:

$$\begin{aligned}\log f_Y(y_t) = &\log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\log((\nu-2)\pi h)\\ &- \frac{\nu+1}{2}\log\left(1+\frac{y_t^2}{(\nu-2)h}\right).\end{aligned}$$

The *score* of f_Y (i.e. the first derivative with respect to h) computed at y_t and standardized by the Fisher information, h^{-2}, is

$$\frac{\partial \log f_Y(y_t)}{\partial h} = \frac{h}{2}\left(\frac{(\nu+1)y_t^2}{(\nu-2)h+y_t^2} - 1\right) = \frac{h}{2}\left((\nu+1)b_t(\nu) - 1\right) \tag{A.1}$$

where we used the notation $b_t(\nu) \equiv y_t^2/((\nu-2)h + y_t^2)$. It is easy to check that the mean of (A.1) is zero. By the properties of Student's t, $b_t(\nu)$ follows the beta distribution with parameters $(1/2, \nu/2)$. The beta distribution with parameters (α, β) characterized by the pdf is

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1],\ \alpha,\beta > 0.$$

where $B(\cdot,\cdot)$ is the beta function. We denote this distribution by $\text{Beta}(\alpha,\beta)$.

APPENDIX B: Functions and Equations

B.1. Derivatives of $l_t(\theta)$

The first three derivatives of $l_t(\theta)$ with respect to β are

$$\frac{\partial l_t(\theta)}{\partial \beta} = \frac{1}{2}h_{\beta t}(\theta)\left[(\nu+1)b_t(\theta) - 1\right],$$

$$
\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \beta^2} &= \frac{1}{2} h_{\beta t}(\theta)^2 \left[(\nu+1) b_t(\theta)(b_t(\theta)-2)+1\right] \\
&\quad + \frac{1}{2} h_{\beta\beta t}(\theta) \left[(\nu+1) b_t(\theta) - 1\right], \qquad \text{(B.1)}\\
\frac{\partial^3 l_t(\theta)}{\partial \beta^3} &= (\nu+1) b_t(\theta)(1-b_t(\theta)) \left[h_{\beta t}(\theta)^3 - \frac{3}{2} h_{\beta t}(\theta) h_{\beta\beta t}(\theta)\right] \qquad \text{(B.2)}\\
&\quad + (\nu+1) h_{\beta t}(\theta)^3 b_t(\theta)(1-b_t(\theta))^2 \\
&\quad - \frac{1}{2}\left(3 h_{\beta t}(\theta) h_{\beta\beta t}(\theta) - 2 h_{\beta t}(\theta)^3 - h_{\beta\beta\beta t}(\theta)\right)\left[(\nu+1) b_t(\theta) - 1\right].
\end{aligned}
$$

Using the fact that

$$
\begin{aligned}
\frac{\partial b_t(\theta)}{\partial \beta} &= -b_t(\theta)(1-b_t(\theta)) h_{\beta t}(\theta), \\
\frac{\partial^2 b_t(\theta)}{\partial \beta^2} &= 2 b_t(\theta)\left(1-b_t(\theta)\right)^2 h_{\beta t}(\theta)^2 - b_t(\theta)\left(1-b_t(\theta)\right) h_{\beta\beta t}(\theta),
\end{aligned}
$$

recursive substitution gives

$$
\begin{aligned}
h_{\beta t}(\theta) &= \sum_{k=1}^{t} \frac{\widehat{h}_{\beta t-k}(\theta)}{\beta + \alpha(\nu+1) b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta + \alpha(\nu+1) b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)}, \\
h_{\beta\beta t}(\theta) &= \sum_{k=1}^{t-1} \frac{\widehat{h}_{\beta\beta t-k}(\theta)}{\beta + \alpha(\nu+1) b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta + \alpha(\nu+1) b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)}, \\
h_{\beta\beta\beta t}(\theta) &= \sum_{k=1}^{t-1} \frac{\widehat{h}_{\beta\beta\beta t-k}(\theta)}{\beta + \alpha(\nu+1) b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)\left(\beta + \alpha(\nu+1) b_{t-j}(\theta)^2\right)}{h_{t-j+1}(\theta)},
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{h}_{\beta t}(\theta) &= 1 \\
\widehat{h}_{\beta\beta t}(\theta) &= 2 h_{\beta t}(\theta)\left[1 - \alpha(\nu+1) b_t(\theta)^2 (1-b_t(\theta)) h_{\beta t}(\theta)\right], \\
\widehat{h}_{\beta\beta\beta t}(\theta) &= 3 h_{\beta\beta t}(\theta) - 6\alpha(\nu+1) b_t(\theta)^2 \left(1-b_t(\theta)\right)^2 h_{\beta t}(\theta)^3 \\
&\quad + 2\alpha(\nu+1) b_t(\theta)^2 \left(1-b_t(\theta)\right) h_{\beta t}(\theta) h_{\beta\beta t}(\theta).
\end{aligned}
$$

With respect to other parameters, we have

$$
\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \delta} &= \frac{1}{2} h_{\delta t}(\theta)\left((\nu+1) b_t(\theta) - 1\right), \\
\frac{\partial l_t(\theta)}{\partial \alpha} &= \frac{1}{2} h_{\alpha t}(\theta)\left((\nu+1) b_t(\theta) - 1\right), \\
\frac{\partial l_t(\theta)}{\partial \gamma} &= \frac{1}{2}\left((\nu+1) b_t(\theta) - 1\right) h_{\gamma t}(\theta) + (\nu+1)\frac{e_t}{e_t^2 + (\nu-2) h_t(\theta)}, \\
\frac{\partial l_t(\theta)}{\partial \nu} &= \frac{1}{2}\left(\psi_0\left(\frac{\nu+1}{2}\right) - \psi_0\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\ln\left(1 + \frac{e_t^2}{(\nu-2) h_t(\theta)}\right) \\
&\quad + \frac{(\nu+1) b_t(\theta) - 1}{2(\nu-2)} + \frac{(\nu+1) b_t(\theta) - 1}{2} h_{\nu t}(\theta),
\end{aligned}
$$

where $\psi_0(\cdot)$ is the digamma function. By recursion, we have

$$h_{\theta_i t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{h}_{\theta_i t-k}(\theta)}{\beta + \alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{(\beta + \alpha(\nu+1)b_{t-j}(\theta)^2)\, h_{t-j}(\theta)}{h_{t-j+1}(\theta)}, \tag{B.3}$$

for $i = 1, 2, \ldots, 6$, where

$$\begin{aligned}
\widehat{h}_{\delta t}(\theta) &= 1/h_t(\theta) \\
\widehat{h}_{\alpha t}(\theta) &= (\nu+1)b_t(\theta) \\
\widehat{h}_{\gamma t}(\theta) &= -2\alpha(\nu+1)(1-b_t(\theta))\frac{e_t}{e_t^2 + (\nu-2)h_t(\theta)} &\text{(B.4)} \\
\widehat{h}_{\nu t}(\theta) &= \alpha b_t(\theta) - \alpha(\nu+1)b_t(\theta)(1-b_t(\theta))(\nu-2)^{-1}. &\text{(B.5)}
\end{aligned}$$

The diagonal elements of $\nabla^2 l_t(\theta)$ are

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \delta^2} &= -\frac{1}{2}h_{\delta t}(\theta)^2((\nu+1)b_t(\theta)-1) - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\delta t}(\theta)^2 \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\delta\delta t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial \alpha^2} &= -\frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\alpha t}(\theta)^2 - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\alpha t}(\theta)^2 \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\alpha\alpha t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial \gamma^2} &= -2(\nu+1)(1-b_t(\theta))h_{\gamma t}(\theta)\frac{e_t}{e_t^2 + (\nu-2)h_t(\theta)} \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta)^2 + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\gamma\gamma t}(\theta) \\
&\quad - \frac{(\nu+1)(1-2b_t(\theta))}{(\nu-2)h_t(\theta) + e_t^2} \\
\frac{\partial^2 l_t(\theta)}{\partial \nu^2} &= 2\psi_1(\nu) + \frac{1}{2(\nu-2)^2} + \frac{1}{2}h_{\nu t}(\theta)^2 - \frac{1}{2}h_{\nu\nu t}(\theta) + b_t(\theta)h_{\nu t}(\theta) \\
&\quad - \frac{\nu+1}{\nu-2}b_t(\theta)(1-b_t(\theta))h_{\nu t}(\theta) - \frac{\nu+1}{2}b_t(\theta)(1-b_t(\theta))h_{\nu t}(\theta)^2 \\
&\quad + \frac{1}{2(\nu-2)^2}b_t(\theta)((\nu+1)b_t(\theta) + (\nu-5)).
\end{aligned}$$

For the cross derivatives, with $i, j = 2, 3, 4$, (i.e. $\theta_i = \alpha$ or β or δ, and $\theta_j = \alpha$ or

β or δ), we have

$$
\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j} &= -\frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\theta_j t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}(\theta)h_{\theta_j t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i\theta_j t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\gamma} &= -\frac{1}{2}(\nu+1)(1-b_t(\theta))h_{\theta_i t}(\theta)\frac{2e_t}{e_t^2+(\nu-2)h_t(\theta)} \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\gamma t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}(\theta)h_{\gamma t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i\gamma t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\nu} &= \frac{1}{2}b_t(\theta)h_{\theta_i t}(\theta) - \frac{\nu+1}{2(\nu-2)}b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta) \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\nu t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}h_{\nu t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\nu\theta_i t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\nu\partial\gamma} &= \frac{1}{2}b_t(\theta)h_{\gamma t}(\theta) - \frac{\nu+1}{2(\nu-2)}b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta) \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta)h_{\nu t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\gamma t}(\theta)h_{\nu t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\nu\gamma t}(\theta) + \frac{1}{\nu-2}\frac{b_t(\theta)}{e_t}((\nu+1)b_t(\theta)-3) \\
&\quad - (\nu+1)\frac{b_t(\theta)}{e_t}(1-b_t(\theta))h_{\nu t}(\theta).
\end{aligned}
$$

By recursion, we have

$$
h_{\theta_i\theta_j t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2)h_{t-j}(\theta)}{h_{t-j+1}(\theta)} \tag{B.6}
$$

for all t and $i, j = 1, 2, \ldots, 6$, where

$$
\begin{aligned}
\widehat{h}_{\alpha\alpha t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)^2 \\
\widehat{h}_{\delta\delta t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)^2, \\
\widehat{h}_{\gamma\gamma t}(\theta) &= -\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))\left[-\frac{2}{e_t^2} + 2b_t(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right)^2\right], \\
\widehat{h}_{\nu\nu t}(\theta) &= \alpha(\nu-2)^{-1}b_t(\theta)h_{\nu t}(\theta)[4(\nu+1)b_t(\theta)^2 - 2(\nu-4)b_t(\theta) - (\nu-2)] \\
&\quad - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta)^2)h_{\nu t}(\theta)^2 \\
&\quad + 2\alpha(\nu-2)^{-2}b_t(\theta)(1-b_t(\theta))(3-(\nu+1)b_t(\theta)), \\
\widehat{h}_{\delta\alpha t}(\theta) &= (\nu+1)b_t(\theta)^2h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)h_{\alpha t}(\theta), \\
\widehat{h}_{\delta\beta t}(\theta) &= h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)h_{\beta t}(\theta), \\
\widehat{h}_{\delta\gamma t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right), \\
\widehat{h}_{\delta\nu t}(\theta) &= \alpha b_t(\theta)^2h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\alpha\beta t}(\theta) &= (\nu+1)b_t(\theta)^2h_{\beta t}(\theta) + h_{\alpha t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)h_{\beta t}(\theta), \\
\widehat{h}_{\alpha\gamma t}(\theta) &= -(\nu+1)b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) + (\nu+1)b_t(\theta)h_{\gamma t}(\theta) \\
&\quad - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right), \\
\widehat{h}_{\alpha\nu t}(\theta) &= b_t(\theta) - (\nu+1)b_t(\theta)(1-b_t(\theta))((\nu-2)^{-1} + h_{\nu t}(\theta)) \\
&\quad + (\nu+1)b_t(\theta)h_{\nu t}(\theta) + \alpha b_t(\theta)^2h_{\alpha t}(\theta) \\
&\quad - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\beta\gamma t}(\theta) &= h_{\gamma t}(\theta) - 2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\beta t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{h}_{\beta\nu t}(\theta) &= h_{\nu t}(\theta) + \alpha b_t(\theta)^2h_{\beta t}(\theta) \\
&\quad - 2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\beta t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\gamma\nu t}(\theta) &= -\alpha b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) + \alpha b_t(\theta)h_{\gamma t}(\theta) \\
&\quad - \alpha(\nu+1)(\nu-2)^{-1}b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta) \\
&\quad + \alpha(\nu+1)(\nu-2)^{-1}(1-2b_t(\theta))b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) \\
&\quad - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\nu t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right).
\end{aligned}
$$

B.2. Definition of $\widetilde{u}_{\theta_i\theta_j t}(\theta_0)$

$$
\begin{aligned}
\widetilde{u}_{\alpha\alpha t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)^2,\\
\widetilde{u}_{\delta\delta t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)^2,\\
\widetilde{u}_{\beta\beta t}(\theta_0) &= 2u^*_{\beta t}(\theta_0)\left(1+\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)\right),\\
\widetilde{u}_{\gamma\gamma t}(\theta_0) &= \alpha_u(\nu_u+1)\left[\frac{1}{(\nu_l-2)\delta_l}+2\left(\frac{2}{(\nu_l-2)\delta_l}+u^*_{\gamma t}(\theta_0)\right)^2\right],\\
\widetilde{u}_{\nu\nu t}(\theta_0) &= \alpha_u(\nu_l-2)^{-1}u^*_{\nu t}(\theta_0)[4(\nu_u+1)+2(\nu_u-4)+(\nu_u-2)]\\
&\quad+2\alpha_u(\nu_u+1)u^*_{\nu t}(\theta_0)^2+2\alpha_u(\nu_l-2)^{-2}(3+(\nu_u+1)),\\
\widetilde{u}_{\delta\alpha t}(\theta_0) &= (\nu_u+1)u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)u^*_{\alpha t}(\theta_0),\\
\widetilde{u}_{\delta\beta t}(\theta_0) &= u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)u^*_{\beta t}(\theta_0),\\
\widetilde{u}_{\delta\gamma t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right),\\
\widetilde{u}_{\delta\nu t}(\theta_0) &= \alpha_u u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)),\\
\widetilde{u}_{\alpha\beta t}(\theta_0) &= (\nu_u+1)u^*_{\beta t}(\theta_0)+u^*_{\alpha t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)u^*_{\beta t}(\theta_0),\\
\widetilde{u}_{\alpha\gamma t}(\theta_0) &= (\nu_u+1)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)+(\nu_u+1)u^*_{\gamma t}(\theta_0)\\
&\quad+2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right),\\
\widetilde{u}_{\alpha\nu t}(\theta_0) &= 1+(\nu_u+1)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0))+(\nu_u+1)u^*_{\nu t}(\theta_0)+\alpha_u u^*_{\alpha t}(\theta_0)\\
&\quad+2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)),\\
\widetilde{u}_{\beta\gamma t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)\\
&\quad+u^*_{\gamma t}(\theta_0),\\
\widetilde{u}_{\beta\nu t}(\theta_0) &= u^*_{\nu t}(\theta_0)+\alpha_u u^*_{\beta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)),\\
\widetilde{u}_{\gamma\nu t}(\theta_0) &= \alpha_u\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)+\alpha_u u^*_{\gamma t}(\theta_0)\\
&\quad+\alpha_u(\nu_u+1)(\nu_u-2)^{-1}u^*_{\gamma t}(\theta_0)\\
&\quad+\alpha_u(\nu_u+1)(\nu_l-2)^{-1}\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)\\
&\quad+2\alpha_u(\nu_u+1)u^*_{\nu t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right).
\end{aligned}
$$

B.3. Definition of $\widehat{u}_{\theta_i\theta_j t}(\theta)$

$$
\begin{aligned}
\widehat{u}_{\delta\delta t}(\theta) &= 0,\\
\widehat{u}_{\alpha\alpha t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)^2\\
\widehat{u}_{\beta\beta t}(\theta) &= 2u_{\beta t}(\theta)\left[1-\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\beta t}(\theta)\right],\\
\widehat{u}_{\gamma\gamma t}(\theta) &= 0,\\
\widehat{u}_{\nu\nu t}(\theta) &= \alpha(\nu-2)^{-1}b_t(\theta)u_{\nu t}(\theta)[4(\nu+1)b_t(\theta)^2-2(\nu-4)b_t(\theta)-(\nu-2)]\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\nu t}(\theta)^2\\
&\quad +2\alpha(\nu-2)^{-2}b_t(\theta)(1-b_t(\theta))(3-(\nu+1)b_t(\theta)),\\
\widehat{u}_{\delta\alpha t}(\theta) &= (\nu+1)b_t(\theta)^2u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)u_{\alpha t}(\theta),\\
\widehat{u}_{\delta\beta t}(\theta) &= u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)u_{\beta t}(\theta),\\
\widehat{u}_{\delta\gamma t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right),\\
\widehat{u}_{\delta\nu t}(\theta) &= \alpha b_t(\theta)^2u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\alpha\beta t}(\theta) &= (\nu+1)b_t(\theta)^2u_{\beta t}(\theta)+u_{\alpha t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)u_{\beta t}(\theta),\\
\widehat{u}_{\alpha\gamma t}(\theta) &= -(\nu+1)b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)+(\nu+1)b_t(\theta)u_{\gamma t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right),\\
\widehat{u}_{\alpha\nu t}(\theta) &= b_t(\theta)-(\nu+1)b_t(\theta)(1-b_t(\theta))((\nu-2)^{-1}+u_{\nu t}(\theta))\\
&\quad +(\nu+1)b_t(\theta)u_{\nu t}(\theta)+\alpha b_t(\theta)^2u_{\alpha t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\beta\gamma t}(\theta) &= u_{\gamma t}(\theta)-2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))u_{\beta t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)\\
\widehat{u}_{\beta\nu t}(\theta) &= u_{\nu t}(\theta)+\alpha b_t(\theta)^2u_{\beta t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))u_{\beta t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\gamma\nu t}(\theta) &= -\alpha b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)+\alpha b_t(\theta)u_{\gamma t}(\theta)\\
&\quad -\alpha(\nu+1)(\nu-2)^{-1}b_t(\theta)(1-b_t(\theta))u_{\gamma t}(\theta)\\
&\quad +\alpha(\nu+1)(\nu-2)^{-1}(1-2b_t(\theta))b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\nu t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right).
\end{aligned}
$$

APPENDIX C: Theorem proofs

Proof for Theorem 1. The proof of this lemma is analogous to Theorem 1 of Nelson (1990). By recursion, we have

$$h_{0t} = \delta_0 \left(1 + \sum_{k=1}^{t-1} \prod_{j=1}^{k} (\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}) \right) + \omega_0 \prod_{j=1}^{t} (\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}) . \tag{C.1}$$

Clearly, $\delta_0 < h_{0t}$ a.s. for all $t \in \mathbb{N}_{>0}$ and any $\theta_0 \in \Theta$. Moreover, (C.1) is absolutely convergent almost surely as $t \to \infty$ if $\mathbb{E}[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})] < 0$ (i.e. $\theta_0 \in \Theta_L$), and otherwise it is divergent [Stout (1974, p. 332) or Theorem 1 of Brandt (1986)].

For all $t \in \mathbb{N}_{>0}$ and $\theta_0 \in \Theta$, the i.i.d. beta random variable b_{0t} is measurable, strictly stationary, and ergodic [Lemma 3.5.8 of Stout (1974)]. Since a finite product of measurable functions is measurable, the product

$$f_k(t, \theta_0) \equiv \prod_{j=1}^{k} (\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}) .$$

is measurable for any t, $\theta_0 \in \Theta$, and $k \in \mathbb{N}_{>0}$. Then $r_n(t, \theta_0) \equiv \sum_{k=1}^{n} f_k(t, \theta_0)$ is measurable for any t, $\theta_0 \in \Theta$, and $n \in \mathbb{N}_{>0}$. [See Royden (1988, p.66-68) for the relevant theorems.] Since $r_n(t, \theta_0)$ is increasing in n,

$$\sup_{n \geq 1} r_n(t, \theta_0) \equiv \sum_{k=1}^{\infty} f_t(k, \theta_0) \tag{C.2}$$

is measurable by Theorem 20 of Royden (1988 p.68) provided that the supremum is finite. The supremum is finite whenever $\theta_0 \in \Theta_L$ by the just established convergence results of h_{0t}. Thus h_{0t} for each $t \in \mathbb{N}$ and $\lim_{t\to\infty} h_{0t}$ are measurable if $\theta_0 \in \Theta_L$. Since b_{0t} is strictly stationary and ergodic for all $\theta_0 \in \Theta$ and h_{0t} is a measurable function of $(b_{0t}, b_{0t-1}, \dots)$ if $\theta_0 \in \Theta_L$, h_{0t} is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974).

The L^p convergence of $(1/h_{0t})_{t\in\mathbb{N}}$ to zero when $\theta_0 \in \Theta_U$ is by dominated convergence, since $0 < 1/h_{0t} \leq 1/\delta_l < \infty$ a.s. for all t and $\theta_0 \in \Theta$. ■

Proof for Theorem 2. By Theorem 1 and Lemma 3, we know that $(h_{0t})_{t\in\mathbb{N}}$, $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$, and $(h_{\theta_i\theta_j t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \dots, 6$. $(b_{0t})_{t\in\mathbb{N}}$ is i.i.d., and so it is also strictly stationary and ergodic. Thus, the desired property holds by Theorem 13.3 of Billingsley (1986) and Theorem 3.5.8 of Stout (1974). ■

Proof for Theorem 3. By Lemmas 16, 8, and 10, the necessary conditions (A.1)-(A.3) in Lemma 1 of Jensen and Rahbek (2004) are satisfied. ■

APPENDIX D: Lemmas

Throughout the following analysis, note that e_t is a function of γ because

$$e_t = \varepsilon_t + (\gamma_0 - \gamma) = \varepsilon_t + g$$

with $g \equiv \gamma_0 - \gamma$. We repeatedly use Lemma 1 to bound several quantities in the subsequent lemmas and obtain the convergence results of Sections 3.

LEMMA 1. *For all $\theta_0, \theta \in \Theta$ and $t \in \mathbb{N}$, we have*

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \begin{cases} 1 & \text{if} \quad |e_t| \geq 1, \\ ((\nu - 2)\delta)^{-1} < \infty & \text{if} \quad |e_t| < 1, \end{cases} \tag{D.1}$$

a.s. If we take the L^p-norm of the LHS quantity, it has the following simple upper-bound;

$$\left\| \frac{e_t}{e_t^2 + (\nu - 2)h_t(\theta)} \right\|_p \leq \sqrt{\frac{1}{(\nu - 2)\delta}} < \infty \tag{D.2}$$

for any $p \geq 1$, $t \in \mathbb{N}$, and $\theta \in \Theta$. Moreover, at $\theta = \theta_0 \in \Theta_U$, the quantity on the LHS of (D.2) tends to zero as $t \to \infty$ for any $p \geq 1$.

Proof. For all $\theta_0, \theta \in \Theta$ and t, we have

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \frac{e_t^2}{e_t^2 + (\nu - 2)h_t(\theta)} \leq 1,$$

if $|e_t| \geq 1$, and

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \frac{1}{(\nu - 2)h_t(\theta)} \leq \frac{1}{(\nu - 2)\delta} < \infty,$$

if $|e_t| < 1$ a.s. This gives (D.1). As the L^p-norms are increasing in p, we have

$$\begin{aligned} \left(\mathbb{E}\left[\left(\frac{e_t}{e_t^2 + (\nu - 2)h_t(\theta)} \right)^p \right] \right)^{1/p} &\leq \sqrt{\frac{1}{(\nu - 2)\delta}} \left(\mathbb{E}\left[\left(\frac{e_t^2}{e_t^2 + (\nu - 2)h_t(\theta)} \right)^p \right] \right)^{1/(2p)} \\ &\leq \sqrt{\frac{1}{(\nu - 2)\delta}} < \infty, \end{aligned}$$

for all t, $p \geq 1$, and $\theta_0, \theta \in \Theta$. This shows (D.2). Finally, using the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y, we obtain for any

$\theta = \theta_0 \in \Theta_U$,

$$\left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_p \leq \left\| \frac{z_t}{z_t^2 + (\nu_0 - 2)} \right\|_{2p} \left\| \frac{1}{\sqrt{h_{0t}}} \right\|_{2p}$$
$$= \left\| \frac{z_t^2}{(z_t^2 + (\nu_0 - 2))^2} \right\|_p^{1/2} \left\| \frac{1}{h_{0t}} \right\|_p^{1/2} \leq \sqrt{\frac{1}{\nu_0 - 2}} \left\| \frac{1}{h_{0t}} \right\|_p^{1/2}$$

and the last quantity on the RHS tends to zero as $t \to \infty$ for any $p \geq 1$ by Theorem 1. ■

The following lemma is used to show that several quantities, especially the derivatives of the log-likelihood, in the subsequent lemmas are bounded in the L^p-norm.

LEMMA 2. *For any $\theta_0 \in \Theta$ and $t \in \mathbb{N}$,*

(i) $h_{\theta_i t}(\theta_0)$ is bounded in L^p for any $p \geq 1$ and $i = 1, \dots, 5$. For instance, we write $\|h_{\beta t}(\theta_0)\|_p \leq H_p(\theta_0) < \infty$ for some $H_p(\theta_0) > 0$.

(ii) $h_{\theta_i \theta_j t}(\theta_0)$ is bounded in L^p for any $p \geq 1$ and $i, j = 1, \dots, 5$. For instance, we write $\|h_{\beta\beta t}(\theta_0)\|_p \leq H_p^\dagger(\theta_0) < \infty$ for some $H_p^\dagger(\theta_0) > 0$.

(iii) $h_{\theta_i \theta_j \theta_k t}(\theta_0)$ is bounded in L^p for any $p \geq 1$ and $i, j, k = 1, \dots, 5$. For instance, we write $\|h_{\beta\beta\beta t}(\theta_0)\|_p \leq H_p^\ddagger(\theta_0) < \infty$ for some $H_p^\ddagger(\theta_0) > 0$.

Proof. (i) For all t and $\theta_0 \in \Theta$, we have

$$0 < \frac{h_{0t}(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2)}{h_{0t+1}} < \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} < 1 \tag{D.3}$$

a.s. because $b_{0t} \in (0, 1)$ a.s. and it is a non-degenerate continuous random variable for each t and $\theta_0 \in \Theta$. Define

$$\left\| \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p \equiv D_p(\theta_0) \in (0, 1)$$

for each $p \geq 1$, t, and $\theta_0 \in \Theta$. We have $D_p(\theta_0) \in (0, 1)$ for each t by (D.3). Note that, for any $k \in \mathbb{N}_{>0}$, $p \geq 1$, and $\theta_0 \in \Theta$, we have

$$\left\| \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p = \prod_{j=1}^{k} \left\| \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p = D_p(\theta_0)^k$$

by the i.i.d. property of $(b_{0t})_{t \in \mathbb{N}}$. Then (B.3) implies that

$$\|h_{\theta_i t}(\theta_0)\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} \left\| \widehat{h}_{\theta_i t-k}(\theta_0) \right\|_{2p} D_{2p}(\theta_0)^k$$

for all t, $\theta_0 \in \Theta$, $p \geq 1$, and $i = 1, \dots, 5$ by the Minkowski inequality and the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y. Since $D_p(\theta_0) \in (0,1)$ for any $p \geq 1$ and $\theta_0 \in \Theta$, $\|h_{\theta_i t}(\theta_0)\|_p$ is bounded in L^p for any $p \geq 1$, t and $\theta \in \Theta$ if so is $\widehat{h}_{\theta_i t-k}(\theta_0)$. For $i = 3$, since $\widehat{h}_{\beta t}(\theta_0) = 1$, we obtain

$$\|h_{\beta t}(\theta_0)\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} D_{2p}(\theta_0)^k \leq \frac{D_{2p}(\theta_0)}{\beta_l(1 - D_{2p}(\theta_0))} \equiv H_p(\theta_0) < \infty. \tag{D.4}$$

for all t, $p \geq 1$, and $\theta_0 \in \Theta$. For $i = 1, 2, 4$, we have

$$\begin{aligned}
&\left|\widehat{h}_{\nu t}(\theta_0)\right| \leq \alpha_0 + \alpha_0(\nu_0 + 1)(\nu_0 - 2)^{-1} < \infty, \\
&\left|\widehat{h}_{\alpha t}(\theta_0)\right| \leq (\nu_0 + 1) < \infty, \quad \left|\widehat{h}_{\delta t}(\theta_0)\right| \leq \delta_0^{-1} < \infty,
\end{aligned}$$

for all t and $\theta_0 \in \Theta$. Thus, we have

$$\begin{aligned}
&\|h_{\nu t}(\theta_0)\|_p \leq (\alpha_u + \alpha_u(\nu_u + 1)(\nu_l - 2)^{-1})H_p(\theta_0) < \infty, \qquad \text{(D.5)} \\
&\|h_{\alpha t}(\theta_0)\|_p \leq (\nu_u + 1)H_p(\theta_0) < \infty, \\
&\|h_{\delta t}(\theta_0)\|_p \leq \frac{D_{2p}(\theta_0)}{\beta_l\delta_l(1 - D_{2p}(\theta_0))} = H_p(\theta_0)/\delta_l < \infty,
\end{aligned}$$

for all t and $\theta_0 \in \Theta$. For $i = 5$, we have

$$\left\|\widehat{h}_{\gamma t}(\theta_0)\right\|_p \leq \frac{2\alpha_0(\nu_0 + 1)}{\sqrt{(\nu_0 - 2)\delta_0}} < \infty$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$ by Lemma 1. Then we have

$$\|h_{\gamma t}(\theta_0)\|_p \leq \frac{2\alpha_u(\nu_u + 1)}{\sqrt{\delta_l(\nu_l - 2)}} H_p(\theta_0) < \infty$$

for any t, $p \geq 1$ and $\theta_0 \in \Theta$.

(ii) By (B.6), we have

$$\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} D_{2p}(\theta_0)^k.$$

for all t, $p \geq 1$, $i, j = 1, \dots, 5$, and $\theta_0 \in \Theta$ by the Minkowski and Hölder inequalities. Thus, $\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p$ is bounded for any $p \geq 1$, t, and $\theta_0 \in \Theta$ if so is $\left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_p$. For instance, we have

$$\left\|\widehat{h}_{\beta\beta t}(\theta_0)\right\|_p \leq 2H_{2p}(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right) < \infty$$

for all t and $\theta_0 \in \Theta$ by Lemma 2 (i). Then we have

$$\begin{aligned}
\|h_{\beta\beta t}(\theta_0)\|_p &\leq \frac{2H_{2p}(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right)}{\beta_l} \sum_{k=1}^{t} D_{2p}(\theta_0)^k \\
&\leq 2H_{2p}(\theta_0)H_p(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right) \equiv H_p^{\dagger}(\theta_0) < \infty
\end{aligned}$$

for all t and $\theta_0 \in \Theta$. Similarly, it is easy to establish that $\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p < \infty$ for all t, $p \geq 1$, $\theta_0 \in \Theta$, and $i, j = 1, 2, 3, 4$ by Lemma 2 (i). For $i = 5$ (or $\theta_i = \gamma$) and $j = 1, \ldots, 5$, first note that

$$\begin{aligned}
&\left|\frac{b_{0t}}{\varepsilon_t^2}\right| = \frac{1}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \leq \frac{1}{(\nu_0 - 2)\delta_0} < \infty, \\
&\left\|b_{0t}\left(\frac{2}{\varepsilon_t} + h_{\gamma t}(\theta_0)\right)\right\|_p \leq 2\left\|\frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}}\right\|_p + \left\|h_{\gamma t}(\theta_0)\right\|_p < \infty,
\end{aligned} \tag{D.6}$$

for all t, $p \geq 1$, and $\theta \in \Theta$ by the Minkowski inequality, Lemma 1, and Lemma 2 (i). Using these properties, it is easy to establish that $\left\|\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right\|_p < \infty$ for any $p \geq 1$, t, and $\theta_0 \in \Theta$ when $i = 5$ and $j = 1, \ldots, 5$.

(iii) Derivations analogous to the above show that the desired property holds for $\left\|h_{\theta_i\theta_j\theta_k t}(\theta_0)\right\|_p$ with $i, j, k = 1, \ldots, 5$ for any t, $p \geq 1$, and $\theta_0 \in \Theta$. For instance, for the case of $i = j = k = 3$, we have

$$\left\|\widehat{h}_{\beta\beta\beta t}(\theta_0)\right\|_p \leq 3H_p^\dagger(\theta_0) + 6\alpha_u(\nu_u + 1)H_{3p}(\theta_0)^3 + 2\alpha_u(\nu_u + 1)H_{2p}(\theta_0)H_{2p}^\dagger(\theta_0) < \infty$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$ by the Minkowski and Hölder inequalities and Lemma 2 (i)(ii). Thus we have

$$\begin{aligned}
&\left\|h_{\beta\beta\beta t}(\theta_0)\right\|_p \\
&\quad\leq \left(3H_p^\dagger(\theta_0) + 6\alpha_u(\nu_u + 1)H_{3p}(\theta_0)^3 + 2\alpha_u(\nu_u + 1)H_{2p}(\theta_0)H_{2p}^\dagger(\theta_0)\right) H_p(\theta_0) \\
&\quad\equiv H_p^\ddagger(\theta_0) < \infty
\end{aligned}$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$. ■

The following lemma is used to show the strict stationarity and ergodicity of $\nabla l_t(\theta_0)$ in Theorem 2.

LEMMA 3. *If $\theta_0 \in \Theta_L$, $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ and $\left(h_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ are strictly stationary and ergodic for $i, j = 1, \ldots, 5$.*

Proof. Note that $0 < (\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2)h_{0t}/h_{0t+1} < 1$ a.s. for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$, and the middle term is strictly stationary and ergodic if $\theta_0 \in \Theta_L$. Then we have

$$\mathbb{E}\left[\log\left(\frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2 h_{0t})}{h_{0t+1}}\right)\right] < 0$$

for all $\theta_0 \in \Theta_L$. By Lemma 1 and Theorem 1, $\left(\widehat{h}_{\theta_i t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic and bounded by some fixed real number a.s. for all $\theta_0 \in \Theta_L$ and

$i = 1, \ldots, 5$. Then we have

$$\mathbb{E}\left[\max\left\{0, \log\left(\frac{\widehat{h}_{\theta_i t}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2}\right)\right\}\right] < \infty.$$

Thus $h_{\theta_i t}(\theta_0)$ is convergent absolutely a.s. for all $t \in \mathbb{N}$, $i = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986). Then $h_{\theta_i t}(\theta_0)$ is measurable for all $t \in \mathbb{N}$, $i = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$. Hence $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i = 1, \ldots, 5$ by Theorem 3.5.8 of Stout (1974).

Then $\left(\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \ldots, 5$. Moreover, using the properties that $b_{0t} \in (0, 1)$ a.s. for all $t \in \mathbb{N}$ and $\max\{0, \log|X|\} \leq |X|$ for any real-valued random variable X, we obtain

$$\mathbb{E}\left[\max\left\{0, \log\left(\frac{\widehat{h}_{\theta_i\theta_j t}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2}\right)\right\}\right] \leq \beta_0^{-1}\mathbb{E}\left[\left|\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right|\right] < \infty$$

for all $t \in \mathbb{N}$, $i, j = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$ by Lemma 2 (i). Then $\left(h_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \ldots, 5$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). ■

Lemma 4 is used to show Lemmas 5 and 9. Note that the condition, $1 \leq p \leq 4$, in Lemma 4 may be relaxed to any $p \geq 1$ if one uses the properties of the beta distribution to express the quantity,

$$\left\|\ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right)\right\|_p = \left\|\ln\left(1 + \frac{z_t^2}{(\nu_0 - 2)}\right)\right\|_p = \left\|\ln(1 - b_{0t})\right\|_p,$$

in terms of polygamma functions. This quantity should be finite for any ν_0 in our parameter space.

LEMMA 4. *For* $1 \leq p \leq 4$, $t \in \mathbb{N}_{>0}$ *and* $\theta_0 \in \Theta$, *we have*

$$\left\|\ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right)\right\|_p < \infty. \tag{D.7}$$

Proof. As the L^p-norm is increasing in $p \geq 1$, it is enough to show (D.7) for $p = 4$. Using the property that $(\ln(1 + x))^4 < 5x$ for all $x > 0$, we have

$$\mathbb{E}\left[\left(\ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right)\right)^4\right] < \mathbb{E}\left[\frac{5\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right] \leq \frac{5}{(\nu_l - 2)} < \infty.$$

■

The preceding results can be used to establish that the elements of $\nabla l_t(\theta_0)$ and $\nabla^2 l_t(\theta_0)$ are bounded in L^p for some p.

LEMMA 5. $\|\partial l_t(\theta_0)/\partial\theta_i\|_p < \infty$ *for* $1 \leq p \leq 4$, $i = 1, \ldots, 5$, $t \in \mathbb{N}_{>0}$, *and any* $\theta_0 \in \Theta$.

Proof. For the derivative with respect to β, we have

$$\left\|\frac{\partial l_t(\theta_0)}{\partial\beta}\right\|_p \leq \frac{\nu_u + 2}{2}\left\|h_{\beta t}(\theta_0)\right\|_p \leq \frac{\nu_u + 2}{2}H_p(\theta_0) < \infty$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$ by Lemma 2 (i). Similar derivations show that $\partial l_t(\theta_0)/\partial\delta$ and $\partial l_t(\theta_0)/\partial\alpha$ are bounded in L^p for all t, $p \geq 1$, and $\theta_0 \in \Theta$ by Lemma 2 (i). $\partial l_t(\theta_0)/\partial\gamma$ is bounded in L^p for all $p \geq 1$, t, and $\theta_0 \in \Theta$ by Lemma 1 and Lemma 2 (i). Finally, $\partial l_t(\theta_0)/\partial\nu$ is bounded in L^4 for all t and $\theta_0 \in \Theta$ by Lemma 2 (i) and Lemma 4. ■

LEMMA 6. $\left\|\partial^2 l_t(\theta_0)/\partial\theta_i\partial\theta_j\right\|_p < \infty$ *for all* $p \geq 1$, $i, j = 1, \ldots, 5$, $t \in \mathbb{N}_{>0}$, *and* $\theta_0 \in \Theta$.

Proof. This can be established by the Minkowski and Hölder inequalities, Lemma 1 and Lemma 2 (i)(ii). For instance, consider $\partial^2 l_t(\theta_0)/\partial\beta^2$. We obtain from (B.1) that

$$\left\|\frac{\partial^2 l_t(\theta_0)}{\partial\beta^2}\right\|_p \leq \frac{2\nu_u + 3}{2}\left\|h_{\beta t}(\theta_0)\right\|_{2p}^2 + \frac{\nu_u + 2}{2}\left\|h_{\beta\beta t}(\theta_0)\right\|_p$$

and the RHS is bounded for all t, $p \geq 1$, and $\theta_0 \in \Theta$ by Lemma 2 (i)(ii). Similar derivations show the desired property for other second derivatives. ■

In order to establish the asymptotic properties of $\nabla L_n(\theta)$ and $\nabla^2 L_n(\theta)$, we define the following new processes

$$u_{\theta_i t}(\theta) = \sum_{k=1}^{t}\frac{\widehat{u}_{\theta_i t-k}(\theta)}{\beta + \alpha(\nu+1)b_{t-k}(\theta)^2}\prod_{j=1}^{k}\frac{\beta + \alpha(\nu+1)b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1)b_{t-j}(\theta)}$$

for $i = 1, \ldots, 5$, where $\widehat{u}_{\theta_i t}(\theta)$ is set to be the limit of $\widehat{h}_{\theta_i t}(\theta)$ when $\theta = \theta_0 \in \Theta_U$ and $t \to \infty$. Thus, we define

$$\widehat{u}_{\nu t}(\theta) = \alpha b_t(\theta) - \alpha(\nu+1)(\nu-2)^{-1}b_t(\theta)(1 - b_t(\theta)),$$
$$\widehat{u}_{\alpha t}(\theta) = (\nu+1)b_t(\theta), \quad \widehat{u}_{\beta t}(\theta) = 1, \quad \widehat{u}_{\delta t}(\theta) = 0, \quad \widehat{u}_{\gamma t}(\theta) = 0.$$

Furthermore, define

$$u_{\theta_i\theta_j t}(\theta) = \sum_{k=1}^{t}\frac{\widehat{u}_{\theta_i\theta_j t-k}(\theta)}{\beta + \alpha(\nu+1)b_{t-k}(\theta)^2}\prod_{j=1}^{k}\frac{\beta + \alpha(\nu+1)b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1)b_{t-j}(\theta)}$$

for $i, j = 1, \ldots, 5$, where $\widehat{u}_{\theta_i\theta_j t}(\theta)$ is set to be the limit of $\widehat{h}_{\theta_i\theta_j t}(\theta)$ when

$\theta = \theta_0 \in \Theta_U$ and $t \to \infty$. They are defined in Appendix B.3. The following lemma establishes some of the useful properties of these processes.

LEMMA 7. *The processes, $(u_{\theta_i t}(\theta))_{t\in\mathbb{N}}$ and $\left(u_{\theta_i\theta_j t}(\theta)\right)_{t\in\mathbb{N}}$, satisfy the following properties.*

(i) $(u_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ *is strictly stationary and ergodic for all* $\theta_0 \in \Theta$ *and* $i = 1, \dots, 5$.

(ii) $u_{\theta_i t}(\theta_0)$ *is bounded in* L^p *for all* $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$, *and* $i = 1, \dots, 5$.

(iii) $0 \leq h_{\theta_i t}(\theta_0) \leq u_{\theta_i t}(\theta_0)$ *for all* $\theta_0 \in \Theta$ *and* $i = 2, 3$ *(i.e.* $\theta_i = \alpha$ *or* β*).*

(iv) Define

$$y^*_{t-k}(\theta) \equiv \prod_{j=1}^{k} \frac{\beta + \alpha(\nu+1)b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1)b_{t-j}(\theta)} - \prod_{j=1}^{k} \frac{(\beta + \alpha(\nu+1)b_{t-j}(\theta)^2)h_{t-j}(\theta)}{h_{t-j+1}(\theta)}$$

for $k, t \in \mathbb{N}$ *and* $\theta, \theta_0 \in \Theta$. *Then* $\|y^*_{t-k}(\theta_0)\|_p \to 0$ *as* $t \to \infty$ *for any* $p \geq 1$, $k \in \mathbb{N}$, *and* $\theta_0 \in \Theta_U$.

(v) $h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0)$ *is bounded in* L^p *for all* $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$ *and* $i = 1, \dots, 5$.

(vi) We have

$$\|h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0)\|_p \to 0 \tag{D.8}$$

as $t \to \infty$ *for any* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i = 1, \dots, 5$.

(vii) We have

$$\left\| \frac{1}{n} \sum_{t=1}^{n} (h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0)) \right\|_p \to 0, \tag{D.9}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} (h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_p \to 0, \tag{D.10}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \right\|_p \to 0, \tag{D.11}$$

as $n \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \dots, 5$.

(viii) $u_{\theta_i\theta_j t}(\theta_0)$ *is strictly stationary ergodic for all* $\theta_0 \in \Theta$ *and* $i, j = 1, \dots, 5$.

(ix) $u_{\theta_i\theta_j t}(\theta_0)$ *is bounded in* L^p *for any* $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$, *and* $i, j = 1, \dots, 5$.

(x) $\left\|h_{\theta_i\theta_j t}(\theta_0) - u_{\theta_i\theta_j t}(\theta_0)\right\|_p < \infty$ *for all* $t \in \mathbb{N}$, $p \geq 1$, $\theta_0 \in \Theta$, *and* $i, j = 1, \ldots, 5$.

(xi) $\left\|h_{\theta_i\theta_j t}(\theta_0) - u_{\theta_i\theta_j t}(\theta_0)\right\|_p \to 0$ *as* $t \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \ldots, 5$.

(xii) $\left\|n^{-1}\sum_{t=1}^{n}\left(h_{\theta_i\theta_j t}(\theta_0) - u_{\theta_i\theta_j t}(\theta_0)\right)\right\|_p \to 0$ *as* $n \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \ldots, 5$.

Proof. **(i) (viii)** By (D.3), we have

$$\mathbb{E}\left[\ln\left(\frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}}\right)\right] < 0.$$

Moreover, using the property that $\ln(x) \leq x - 1$ for all $x > 0$, we have

$$\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i}|\}\right] \leq \mathbb{E}\left[\max\{0, |\widetilde{u}_{\theta_i}|\}\right] = \mathbb{E}\left[|\widetilde{u}_{\theta_i}|\right] < \infty$$

for $i = 1, \ldots, 5$. Then $(u_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta$ by Theorem 1 of Brandt (1986). Likewise, we can show that $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j t}(\theta_0)|\}\right] < \infty$ and $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j\theta_m t}(\theta_0)|\}\right] < \infty$ for all t, $\theta_0 \in \Theta$, and $i, j, m = 1, \ldots, 5$. Then we can deduce that $(u^*_{\theta_i\theta_j t}(\theta_0))_{t\in\mathbb{N}}$ and $(u^*_{\theta_i\theta_j\theta_m t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for any $\theta_0 \in \Theta$ and $i, j, m = 1, \ldots, 5$ by Theorem 1 of Brandt (1986).

(ii) The proof for $i = 4, 5$ (i.e. $\theta_i = \delta$ or γ) is trivial as $\widehat{u}_{\delta t}(\theta) = \widehat{u}_{\gamma t}(\theta) = 0$ for all t and $\theta \in \Theta$. Recalling (D.3) and $D_p(\theta_0)$ defined in in Lemma 2 (i), we obtain

$$\begin{aligned}
\|u_{\alpha t}(\theta_0)\|_p &\leq \frac{\nu_u+1}{\beta_l}\sum_{k=1}^{\infty} D_p(\theta_0)^k = \frac{(\nu_u+1)D_p(\theta_0)}{\beta_l(1-D_p(\theta_0))} < \infty, \\
\|u_{\beta t}(\theta_0)\|_p &\leq \frac{1}{\beta_l}\sum_{k=1}^{\infty} D_p(\theta_0)^k = \frac{D_p(\theta_0)}{\beta_l(1-D_p(\theta_0))} < \infty, \\
\|u_{\nu t}(\theta_0)\|_p &\leq \frac{\alpha_u + \alpha_u(\nu_u+1)(\nu_l-2)^{-1}}{\beta_l}\sum_{k=1}^{\infty} D_p(\theta_0)^k \\
&\leq \frac{(\alpha_u + \alpha_u(\nu_u+1)(\nu_l-2)^{-1})D_p(\theta_0)}{\beta_l(1-D_p(\theta_0))} < \infty.
\end{aligned} \tag{D.12}$$

for all t, $p \geq 1$, and $\theta_0 \in \Theta$.

(iii) This is by (D.3) and the fact that $\widehat{u}_{\theta_i t}(\theta_0) = \widehat{h}_{\theta_i t}(\theta_0)$ for all t, $\theta_0 \in \Theta$, and $i = 2, 3$ (i.e. $\theta_i = \alpha$ or β).

(iv) By Theorem 1 and (D.3), we have

$$0 < \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} - \frac{(\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2)h_{0t-j}}{h_{0t-j+1}} \to 0$$

a.s. as $t \to \infty$ for any $j \in \mathbb{N}$ and $\theta_0 \in \Theta_U$. Thus $0 \leq y^*_{t-k}(\theta_0) \to 0$ a.s. as $t \to \infty$

for any $k \in \mathbb{N}$ and all $\theta_0 \in \Theta_U$. Moreover, $y_t^*(\theta)^p \leq 1$ for any $p \geq 1$, $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Thus $\|y_{t-k}^*(\theta_0)\|_p \to 0$ as $t \to \infty$ for any $k \in \mathbb{N}$, $p \geq 1$, and $\theta_0 \in \Theta_U$ by dominated convergence.

(v) This is by the Minkowski inequality, Lemma 2 (i), and Lemma 7 (ii).

(vi) We prove (D.8) for $p = 1$ first. For any $t_0 < t$ and $\theta_0 \in \Theta_U$, we have

$$
\begin{aligned}
&\mathbb{E}\left[|u_{\alpha t}(\theta_0) - h_{\alpha t}(\theta_0)|\right] \\
&\quad= \mathbb{E}\left[\left|\sum_{k=1}^{\infty} \frac{(\nu_0+1)b_{0t-k}}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}}\right.\right. \\
&\qquad\left.\left. - \sum_{k=1}^{t} \frac{(\nu_0+1)b_{0t-k}}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{h_{0t-j}(\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2)}{h_{0t-j+1}}\right|\right] \\
&\quad\leq \sum_{k=1}^{t_0} \mathbb{E}\left[\left|\frac{(\nu_0+1)b_{0t-k}}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} y_{t-k}^*(\theta_0)\right|\right] \\
&\qquad + \sum_{k=t_0+1}^{\infty} \mathbb{E}\left[\left|\frac{(\nu_0+1)b_{0t-k}}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}}\right|\right] \\
&\quad\leq \frac{\nu_u+1}{\beta_l} \sum_{k=1}^{t_0} \mathbb{E}\left[|y_{t-k}^*(\theta_0)|\right] + \frac{\nu_u+1}{\beta_l} \frac{D_1(\theta_0)^{t_0+1}}{1 - D_1(\theta_0)}
\end{aligned}
$$

by the triangle inequality and (D.3). The first term in the final line tends to zero as $t \to \infty$ for any $t_0 < t$ and $\theta_0 \in \Theta_U$ by Lemma 7 (iv). As the choice of $t_0 < t$ was arbitrary and $D_1(\theta_0) \in (0,1)$ for any $\theta_0 \in \Theta$, the second term in the final line tends to zero as $t_0 \to \infty$. Thus $\|u_{\alpha t}(\theta_0) - h_{\alpha t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. Analogous derivations show that $\|u_{\beta t}(\theta_0) - h_{\beta t}(\theta_0)\|_1 \to 0$ and $\|u_{\nu t}(\theta_0) - h_{\nu t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. For $i = 4$ (i.e. $\theta_i = \delta$), note that, for any $k \in \mathbb{N}$, we have

$$0 \leq \left(\frac{1}{(\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right)^p \to 0$$

a.s. as $t \to \infty$ for any $p \geq 1$ and $\theta_0 \in \Theta_U$ by Theorem 1. The term in the middle is also bounded above by $1/(\beta_l\delta_l)^p$ a.s. Thus, by dominated convergence,

$$\left\|\frac{1}{(\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right\|_p \to 0 \tag{D.13}$$

as $t \to \infty$ for any $p \geq 1$, $k \in \mathbb{N}$, and all $\theta_0 \in \Theta_U$. Then, for any arbitrary $t_0 < t$,

$$
\begin{aligned}
0 \leq \|h_{\delta t}(\theta_0) - u_{\delta t}(\theta_0)\|_1 &\leq \sum_{k=1}^{t_0} \left\|\frac{1}{(\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right\|_2 D_2(\theta_0)^k \\
&\quad + \frac{D_2(\theta_0)^{t_0+1}}{\beta_0\delta_0(1 - D_2(\theta_0))}.
\end{aligned}
$$

By (D.13), the first term on the RHS tends to zero as $t \to \infty$ for any $t_0 < t$ and $\theta_0 \in \Theta_U$. The second term on the RHS also tends to zero as $t_0 \to \infty$. Hence $\|h_{\delta t}(\theta_0) - u_{\delta t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. For $i = 5$ (i.e. $\theta_i = \gamma$), note that

$$0 \leq \mathbb{E}\left[|\widehat{h}_{\gamma t-k}(\theta_0)|\right] \leq 2\alpha_0(\nu_0 + 1)\mathbb{E}\left[\frac{|\varepsilon_t|}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}}\right] \to 0$$

as $t \to \infty$ for all $\theta_0 \in \Theta_U$ by Lemma 1. Moreover, $\widehat{h}_{\gamma t}(\theta_0)$ is bounded in L^p for any $p \geq 1$, $\theta_0 \in \Theta$, and t by Lemma 2 (i). Thus, we can show that $\|h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ by derivations similar to the $i = 4$ (i.e. $\theta_i = \delta$) case.

By these properties of L^1 convergence to zero and the uniform integrability established in Lemma 7 (v), we have (D.8) for any $p \geq 1$, $\theta_0 \in \Theta_U$, and $i = 1, \ldots, 5$.

(vii) (D.9) is by the Minkowski inequality and Lemma 7 (vi). We also have

$$\begin{aligned}
&\left\| \frac{1}{n}\sum_{t=1}^{n} \left(h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)\right) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_p \\
&\qquad \leq \frac{1}{n}\sum_{t=1}^{n} \|h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)\|_{2p} \left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_{2p} \to 0,
\end{aligned}$$

as $n \to \infty$ for any $\theta_0 \in \Theta_U$ by Lemma 7 (vi) and Lemma 1. This establishes (D.10). Next we show (D.11). For any $p \geq 1$, $\theta_0 \in \Theta_U$, and $i, j = 1, \ldots, 5$, we have

$$\begin{aligned}
&\left\|u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0) - h_{\theta_i t}(\theta_0)h_{\theta_j t}(\theta_0)\right\|_p \\
&\leq \left\|u_{\theta_i t}(\theta_0)\left(u_{\theta_j t}(\theta_0) - h_{\theta_j t}(\theta_0)\right)\right\|_p + \left\|h_{\theta_j t}(\theta_0)(u_{\theta_i t}(\theta_0) - h_{\theta_i t}(\theta_0))\right\|_p \\
&\leq \|u_{\theta_i t}(\theta_0)\|_{2p}\left\|u_{\theta_j t}(\theta_0) - h_{\theta_j t}(\theta_0)\right\|_{2p} + \left\|h_{\theta_j t}(\theta_0)\right\|_{2p}\|u_{\theta_i t}(\theta_0) - h_{\theta_i t}(\theta_0)\|_{2p} \\
&\to 0
\end{aligned} \tag{D.14}$$

as $t \to \infty$ by Lemma 2 (i) and Lemma 7 (ii)(vi). Thus we obtain

$$\begin{aligned}
&\left\| \frac{1}{n}\sum_{t=1}^{n} \left(h_{\theta_i t}(\theta_0)h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\right) \right\|_p \\
&\qquad \leq \frac{1}{n}\sum_{t=1}^{n} \left\|\left(h_{\theta_i t}(\theta_0)h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\right)\right\|_p \to 0
\end{aligned}$$

as $n \to \infty$ for all $\theta_0 \in \Theta_U$, $p \geq 1$, and $i, j = 1, \ldots, 5$. This shows (D.11).

(ix) The proof is analogous to Lemma 2 (i)(ii), and thus omitted.

(x) This is by the Minkowski inequality, Lemma 2 (ii), and Lemma 7 (ix).

(xi) For any $t_0 < t$, we have

$$
\begin{aligned}
&\left\|h_{\theta_i\theta_j t}(\theta_0) - u_{\theta_i\theta_j t}(\theta_0)\right\|_p \\
&\quad \le \sum_{k=1}^{t_0} \left\| \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} y_{t-k}^*(\theta_0) \right\|_p \\
&\qquad + \sum_{k=1}^{t_0} \left\| \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta_0) - \widehat{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \right\|_p \\
&\qquad + \sum_{k=t_0+1}^{t} \left\| \frac{\widehat{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \right\|_p \qquad \text{(D.15)} \\
&\quad \le \beta_l^{-1} \sum_{k=1}^{t_0} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} \left\|y_{t-k}^*(\theta_0)\right\|_{2p} \\
&\qquad + \beta_l^{-1} \sum_{k=1}^{t_0} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0) - \widehat{u}_{\theta_i\theta_j t-k}(\theta_l)\right\|_{2p} D_{2p}(\theta_0)^k \\
&\qquad + \beta_0^{-1} \sup_{i,j} \left\|\widehat{u}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} \frac{D_{2p}(\theta_0)^{t_0+1}}{1 - D_{2p}(\theta_0)}
\end{aligned}
$$

for all t and $\theta_0 \in \Theta$. The second inequality used the property that $\widehat{u}_{\theta_i\theta_j t}(\theta_0)$ is bounded in L^p and strictly stationary for all $p \ge 1$, $\theta_0 \in \Theta$, and $i, j = 1, \ldots, 5$. Note that we have

$$\left\|\widehat{h}_{\theta_i\theta_j t}(\theta_0) - \widehat{u}_{\theta_i\theta_j t}(\theta_0)\right\|_p \to 0 \qquad \text{(D.16)}$$

as $t \to \infty$ for any $p \ge 1$, $\theta_0 \in \Theta_U$, and $i, j = 1, \ldots, 5$ by (D.14), Lemma 7 (vi), and Lemma 1. Then, by (D.16), Lemma 2 (ii), Lemma 7 (iv)(ix), and the property that $D_p(\theta_0) \in (0,1)$ for all $p \ge 1$ and $\theta_0 \in \Theta$, the terms after the second inequality of (D.15) tends to zero as $t \to \infty$ and $t_0 \to \infty$ (since the choice of $t_0 < t$ was arbitrary) for any $\theta_0 \in \Theta_U$, $p \ge 1$, and $i, j = 1, \ldots, 5$.

(xii) This is by Lemma 7 (xi) and derivations analogous to the proof for (D.9) in Lemma 7 (vii). ∎

Lemma 7 is used in the following lemma, which is used to show the asymptotic property of $\nabla_\theta L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0^*)$.

LEMMA 8. *Assume that $\theta_0 \in \Theta_L$. Then*

$$\frac{1}{n} \sum_{t=1}^{n} \nabla_\theta^2 l_t(\theta_0) \xrightarrow{P} \mathbb{E}\left[\nabla_\theta^2 l_t(\theta_0)\right] \equiv Q(\theta_0), \qquad \text{(D.17)}$$

where $Q(\theta_0)$ is a constant symmetric matrix given θ_0. Moreover, if $\theta_0 \in \Theta_U$, then

$$\frac{1}{n}\sum_{t=1}^{n} \nabla^2_{\theta^*} l_t(\theta_0) \xrightarrow{P} \mathbb{E}\left[\nabla^2_{\theta^*} l_t(\theta_0)\right] \equiv Q^*(\theta_0) \tag{D.18}$$

for all $\theta_0 \in \Theta_U$, where $Q^*(\theta_0)$ is a constant symmetric matrix given θ_0.

Proof. For all $\theta \in \Theta_L$, $\nabla^2_\theta l_t(\theta_0)$ is strictly stationary and ergodic by Theorem 2. Moreover, $\mathbb{E}\left[|\partial^2 l_t(\theta_0)/\partial\theta_i\partial\theta_j|\right] < \infty$ for all $\theta_0 \in \Theta$ and $i, j = 1, \ldots, 5$ by Lemma 6. Thus (D.17) holds for all $\theta_0 \in \Theta_L$ by replacing the almost sure convergence of Theorem 3.5.7 of Stout (1974) by convergence in probability. For $\theta_0 \in \Theta_U$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\beta^2} &= \frac{1}{2n}\sum_{t=1}^{n}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]\left(h_{\beta t}(\theta_0)^2 - u_{\beta t}(\theta_0)^2\right) \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\beta\beta t}(\theta_0) - u_{\beta\beta t}(\theta_0)\right) \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]u_{\beta t}(\theta_0)^2 \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)u_{\beta\beta t}(\theta_0) \\
&\xrightarrow{P} \frac{1}{2}\mathbb{E}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]\mathbb{E}\left[u_{\beta t}(\theta_0)^2\right] \\
&= \frac{1}{2}\left(\frac{3}{\nu_0+3} - 1\right)\mathbb{E}\left[u_{\beta t}(\theta_0)^2\right] < 0
\end{aligned}$$

as $n \to \infty$ by Lemma 7 (i)(vii)(viii)(xi)(xii) and the property that b_{0t} for each $t \in \mathbb{N}$ is i.i.d. with the distribution Beta$(1/2, \nu_0/2)$. Similarly, we can use the results of Lemma 7 to establish the desired convergence of other diagonal and off-diagonal elements of (D.18). ■

The following lemma verifies that the limiting distribution of $\nabla_\theta L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0)$ in Lemma 10 have well-defined variances.

LEMMA 9. $\mathbb{E}\left[|(\partial l_t(\theta_0)/\partial\theta_i)(\partial l_t(\theta_0)/\partial\theta_j)|\,\right] < \infty$ *for all $t \in \mathbb{N}$, $i, j = 1, \ldots, 5$, and $\theta_0 \in \Theta$.*

Proof. Using the property that $\|X^2\|_1 = \|X\|_2^2$ for any random variable X, we know that $\mathbb{E}\left[|\partial l_t(\theta_0)/\partial\theta_i|^2\right]$ is bounded for all $\theta_0 \in \Theta$ and $i = 1, \ldots, 5$ by Lemma 5. For $i \neq j$, $(\partial l_t(\theta_0)/\partial\theta_i)(\partial l_t(\theta_0)/\partial\theta_j)$ are also bounded in L^1 since

$$\left\|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\right\|_1 \leq \left\|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\right\|_2\left\|\frac{\partial l_t(\theta_0)}{\partial\theta_j}\right\|_2 < \infty$$

for all $i, j = 1, \ldots, 5$ and $i \neq j$ by the Hölder inequality and Lemma 5. ■

The following proposition is used to establish Lemma 10.

PROPOSITION 1. Let $(X_t)_{t=1}^{\infty}$ be a sequence of random variables satisfying $n^{-1}\sum_{t=1}^{n} X_t \xrightarrow{P} 0$ as $n \to \infty$. If $n^{-1}\sum_{t=1}^{n} \mathbb{E}[X_t]$ converges as $n \to \infty$, then its limit is zero.

Proof. We can find a subsequence $(X_{t_k})_{k=1}^{\infty}$ such that $n^{-1}\sum_{k=1}^{n} X_{t_k} \to 0$ a.s. as $n \to \infty$. Then $n^{-1}\sum_{k=1}^{n} \mathbb{E}[X_{t_k}] \to 0$ as $n \to \infty$, and since $n^{-1}\sum_{t=1}^{n} \mathbb{E}[X_t]$ is convergent, its limit must be zero. ■

We are now ready to show that the asymptotic distribution of $\nabla_{\theta} L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0)$ are normal with a well-defined covariance matrices.

LEMMA 10. *For all* $\theta_0 \in \Theta_L$,

$$R(\theta_0)^{-1/2}\sqrt{n}\nabla_{\theta} L_n(\theta_0) \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as} \quad n \to \infty, \tag{D.19}$$

where $R(\theta_0) \equiv \mathbb{E}[\nabla_{\theta} l_t(\theta_0)\nabla_{\theta} l_t(\theta_0)^{\top}]$. *Moreover, for all* $\theta_0 \in \Theta_U$,

$$R^*(\theta_0)^{-1/2}\sqrt{n}\nabla_{\theta^*} L_n(\theta_0) \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as} \quad n \to \infty, \tag{D.20}$$

where $R^*(\theta_0) \equiv \mathbb{E}[\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^{\top}]$.

Proof. We first verify that $(\nabla_{\theta} l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_{\theta^*} l_t(\theta_0))_{t\in\mathbb{N}}$ are martingale difference sequences (MD). Since $(b_{0t})_{t\in\mathbb{N}}$ is i.i.d. with the distribution, Beta$(1/2, \nu_0/2)$, for each t, we have

$$\mathbb{E}\left[\left.\frac{\partial l_t(\theta_0)}{\partial \delta}\right| \mathcal{F}_{t-1}\right] = \mathbb{E}\left[\left.\frac{\partial l_t(\theta_0)}{\partial \alpha}\right| \mathcal{F}_{t-1}\right] = \left[\left.\frac{\partial l_t(\theta_0)}{\partial \beta}\right| \mathcal{F}_{t-1}\right] = 0,$$

for all t. Moreover, we have

$$\mathbb{E}\left[\left.\ln\left(1 + \frac{z_t^2}{(\nu_0 - 2)}\right)\right| \mathcal{F}_{t-1}\right] = -\mathbb{E}\left[\ln(1 - b_{0t})\right| \mathcal{F}_{t-1}] = \psi_0\left(\frac{\nu_0 + 1}{2}\right) - \psi_0\left(\frac{\nu_0}{2}\right),$$

for all t by the properties of the beta distribution. Thus $\mathbb{E}\left[\partial l_t(\theta_0)/\partial \nu | \mathcal{F}_{t-1}\right] = 0$. Finally, we have $\mathbb{E}\left[\partial l_t(\theta_0)/\partial \gamma \right| \mathcal{F}_{t-1}] = 0$ if

$$\mathbb{E}\left[\left.\frac{z_t}{z_t^2 + (\nu_0 - 2)}\right| \mathcal{F}_{t-1}\right] = 0. \tag{D.21}$$

Computing the integral directly, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\frac{z_t}{z_t^2+(\nu_0-2)}\middle|\mathcal{F}_{t-1}\right] &\propto \int_{-\infty}^{\infty}\frac{z}{z^2+(\nu_0-2)}\left(1+\frac{z^2}{\nu_0-2}\right)^{-\frac{\nu_0+1}{2}}dz \\
&= \int_{-\pi/2}^{\pi/2}\frac{\tan x}{\tan^2 x+1}\left(1+\tan^2 x\right)^{-\frac{\nu_0+1}{2}}\sec^2 x\, dx \\
&\propto \int_{-\pi/2}^{\pi/2}\sin x\,(\cos x)^{\nu_0}\, dx \\
&= 0,
\end{aligned}
$$

for all $t \in \mathbb{N}$, where the second line is by the change of variable, $z/\sqrt{\nu_0-2} = \tan x$, the third line is by basic trigonometric identities, and the last line is by the fact that the integrand is an odd function for any $\nu_0 \in \mathbb{R}$. Thus (D.21) holds. Then, since $\nabla_\theta l_t(\theta_0)$ and $\nabla_{\theta^*} l_t(\theta_0)$ are integrable for all $\theta_0 \in \Theta_L$ and $\theta_0 \in \Theta_U$, respectively, by Lemma 5, $(\nabla l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_{\theta^*} l_t(\theta_0))_{t\in\mathbb{N}}$ are MDs.

If $\theta_0 \in \Theta_L$, $\nabla_\theta l_t(\theta_0)$ is a strictly stationary and ergodic martingale difference with finite unconditional second moment by Theorem 2 and Lemma 9. Thus (D.19) holds at $\theta = \theta_0$ by the central limit theorem for stationary ergodic martingales [Theorem 6.11 of Varadhan (2001, p.144)].

For $\theta_0 \in \Theta_U$, we aim to show that

$$
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^\top\right] \to R^*(\theta_0), \tag{D.22}
$$

$$
\mathbb{E}\left[\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\frac{\partial l_t(\theta_0)}{\partial\theta_k}\right] < \infty \text{ for all } t, \text{ and} \tag{D.23}
$$

$$
\frac{1}{n}\sum_{t=1}^{n}\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^\top \xrightarrow{P} R^*(\theta_0), \tag{D.24}
$$

where $i, j, k = 1, 2, 3$, convergence is as $n \to \infty$, and $R^*(\theta_0)$ is a deterministic and finite positive definite matrix for each $\theta_0 \in \Theta_U$. (D.22)-(D.24) imply that (D.20) holds for $\theta_0 \in \Theta_U$ by Proposition 7.9 of Hamilton (1994, p.194). (Note that the proofs for (D.22) and (D.23) presented below holds for all $\theta_0 \in \Theta$, whereas the proof for (D.24) holds only for $\theta_0 \in \Theta_U$.)

By the integrability of $\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^\top$ shown in Lemma 9, $\mathbb{E}\left[\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^\top\right] \equiv R_t^*(\theta_0)$ is a finite positive definite matrix for each $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Since $(R_t^*(\theta_0))_{t\in\mathbb{N}}$ is a deterministic sequence of real matrices, its sample average, $n^{-1}\sum_{t=1}^{n} R_t^*(\theta_0)$, converges to some constant positive definite matrix $R^*(\theta_0)$ as $n \to \infty$. (This convergence is verified easily by considering a special case of the law of large numbers where the sequence of i.i.d. random variables are replaced by a deterministic sequence.) Thus (D.22) holds for all

$\theta_0 \in \Theta$.

(D.23) holds if

$$\left\| \frac{\partial l_t(\theta_0)}{\partial \theta_i} \right\|_3 \left\| \frac{\partial l_t(\theta_0)}{\partial \theta_j} \right\|_3 \left\| \frac{\partial l_t(\theta_0)}{\partial \theta_k} \right\|_3 < \infty \tag{D.25}$$

for $i, j, k = 1, 2, 3$ and all $t \in \mathbb{N}$, since

$$|\mathbb{E}[XYZ]| \leq \|XYZ\|_1 \leq \|X\|_3 \|Y\|_3 \|Z\|_3$$

for any random variables X, Y, and Z. (D.25) holds if $\nabla_{\theta^*} l_t(\theta_0)$ is bounded in L^3. By Lemma 5, (D.25) holds for all $\theta_0 \in \Theta$. Thus, we have (D.23) for all $\theta_0 \in \Theta$.

Finally, we show (D.24) for $\theta_0 \in \Theta_U$. If we can show that $n^{-1} \sum_{t=1}^n \nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top$ converges in probability to *some* constant positive definite matrix as $n \to \infty$, then the limiting quantity must be the same as the RHS of (D.22) by Proposition 1. Thus there is no need to verify that the limit of (D.22) and (D.24) are the same. We deal with the diagonal elements first. For all $\theta_0 \in \Theta_U$ and $i = 2, 3$, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial l_t(\theta_0)}{\partial \theta_i} \right)^2 &= \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 (h_{\theta_i t}(\theta_0)^2 - u_{\theta_i t}(\theta_0)^2) \\ &\quad + \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 u_{\theta_i t}(\theta_0)^2, \end{aligned}$$

where the RHS converges in probability to $((\nu_0+1)^2/4)\mathrm{Var}(b_{0t})\mathbb{E}[u_{it}(\theta_0)^2] < \infty$ by Lemma 7 (i)(ii)(vii). (Note also that b_{0t} and $u_{it}(\theta_0)$ are independent for all t.)

Next, we have

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \left(\frac{\partial l_t(\theta_0)}{\partial \nu} \right)^2 \\ &= \frac{1}{n} \sum_{t=1}^n \left[\frac{(\nu_0 + 1) b_{0t} - 1}{2(\nu_0 - 2)} - \frac{1}{2} \ln \left(1 + \frac{z_t^2}{\nu_0 - 2} \right) + \frac{1}{2} \left(\psi_0 \left(\frac{\nu_0 + 1}{2} \right) - \psi_0 \left(\frac{\nu_0}{2} \right) \right) \right]^2 \\ &+ \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 \left(h_{\nu t}(\theta_0)^2 - u_{\nu t}(\theta_0)^2 \right) \\ &+ \frac{2}{n} \sum_{t=1}^n \left\{ \left[\frac{(\nu_0 + 1) b_{0t} - 1}{2(\nu_0 - 2)} - \frac{1}{2} \ln \left(1 + \frac{z_t^2}{\nu_0 - 2} \right) + \frac{1}{2} \left(\psi_0 \left(\frac{\nu_0 + 1}{2} \right) - \psi_0 \left(\frac{\nu_0}{2} \right) \right) \right] \right. \\ &\qquad \left. \times \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right) (h_{\nu t}(\theta_0) - u_{\nu t}(\theta_0)) \right\} \\ &+ \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 u_{\nu t}(\theta_0)^2 \end{aligned}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left\{\left[\frac{(\nu_0+1)b_{0t}-1}{2(\nu_0-2)}-\frac{1}{2}\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{1}{2}\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)\right)\right]\right.$$
$$\left.\times\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)u_{\nu t}(\theta_0)\right\}$$

The summand of the first, fourth, and fifth terms are stationary and ergodic by Lemma 7 (i). By (D.9) and (D.11) of Lemma 7 (vii), and by the properties that $u_{\nu t}(\theta_0)$ and $h_{\nu t}(\theta_0)$ are independent of b_{0t} and z_t, the second and third terms converge to zero in L^1 for all $\theta_0\in\Theta_U$. Then we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial\nu}\right)^2 \xrightarrow{P} &\operatorname{Var}\left(\frac{(\nu_0+1)b_{0t}}{2(\nu_0-2)}-\frac{1}{2}\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)\right)\\ &+\frac{(\nu_0+1)^2}{4}\operatorname{Var}(b_{0t})\mathbb{E}[u_{\nu t}(\theta_0)^2]\\ &+\frac{\nu_0+1}{2}\operatorname{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right),b_{0t}\right)\mathbb{E}[u_{\nu t}(\theta_0)]\\ &<\infty\end{aligned}$$

as $n\to\infty$ for all $\theta_0\in\Theta_U$.

Next, we consider the off-diagonal elements of (D.24). We have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j} &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2\left(h_{\theta_i t}(\theta_0)h_{\theta_j t}(\theta_0)-u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\right)\\ &\quad+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2 u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\\ &\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\operatorname{Var}(b_{0t})\mathbb{E}[u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)]<\infty\end{aligned}$$

for all $\theta_0\in\Theta_U$, $i,j=2,3$, and $i\neq j$ by Lemma 7 (i)(ii)(vii). Similarly, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\beta}\frac{\partial l_t(\theta_0)}{\partial\nu} &= \frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\beta t}(\theta_0)h_{\nu t}(\theta_0)-u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\right)\\ &+\frac{1}{4n}\sum_{t=1}^{n}\left\{\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right)\right.\\ &\qquad\left.\times\left((\nu_0+1)b_{0t}-1\right)\left(h_{\beta t}(\theta_0)-u_{\beta t}(\theta_0)\right)\right\}\\ &+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\end{aligned}$$

$$
+\frac{1}{4n}\sum_{t=1}^{n}\left\{\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right)\right.
$$

$$
\left.\times\left((\nu_0+1)b_{0t}-1\right)u_{\beta t}(\theta_0)\right\}
$$

$$
\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\mathrm{Var}(b_{0t})\mathbb{E}\left[u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\right]
$$

$$
+\frac{\nu_0+1}{4}\mathrm{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right),b_{0t}\right)\mathbb{E}\left[u_{\beta t}(\theta_0)\right]<\infty.
$$

for all $\theta_0 \in \Theta_U$ by Lemma 7 (i)(ii)(vii). Analogous derivations show that

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\alpha}\frac{\partial l_t(\theta_0)}{\partial\nu}\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\mathrm{Var}(b_{0t})\mathbb{E}\left[u_{\alpha t}(\theta_0)u_{\nu t}(\theta_0)\right]
$$

$$
+\frac{\nu_0+1}{4}\mathrm{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right),b_{0t}\right)\mathbb{E}\left[u_{\alpha t}(\theta_0)\right].
$$

for all $\theta_0 \in \Theta_U$. This completes the proof of (D.24) for $\theta_0 \in \Theta_U$. ■

In the next lemma, we show that, if $\theta_0 \in \Theta_U$, the joint log-likelihood function is asymptotically flat in the δ and γ dimensions, so that the consistency and asymptotic normality of MLE do not hold for these parameters when $\theta_0 \in \Theta_U$.

LEMMA 11. *For $i = 4, 5$ and $j = 1, \ldots, 5$, we have*

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\xrightarrow{P}0.
$$

when $\theta_0 \in \Theta_U$.

Proof. For all $\theta_0 \in \Theta_U$ and $i = 4$ (i.e. $\theta_i = \delta$), we have

$$
\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial\delta}\right)^2=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2\left(h_{\delta t}(\theta_0)^2-u_{\delta t}(\theta_0)^2\right)
$$

$$
+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2u_{\delta t}(\theta_0)^2,
$$

where the RHS converges in probability to $((\nu_0+1)^2/4)\mathrm{Var}(b_{0t})\mathbb{E}[u_{\delta t}(\theta_0)^2] = 0$ by Lemma 7 (i)(ii)(vii). Next, we have

$$
\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial\gamma}\right)^2=\frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\gamma t}(\theta_0)^2-u_{\gamma t}(\theta_0)^2\right)
$$

$$
+\frac{\nu_0+1}{n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\gamma t}(\theta_0)-u_{\gamma t}(\theta_0)\right)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}
$$

$$
+\frac{(\nu_0+1)^2}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{(\varepsilon_t^2+(\nu_0-2)h_{0t})^2}.
$$

The first two terms on the RHS converges in probability to zero by (D.10) and (D.11) of Lemma 7 (vii). The third term converges in L^1 to zero because

$$0 \leq \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{(\varepsilon_t^2 + (\nu_0 - 2)h_{0t})^2} \right\|_1 \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_2^2 \to 0$$

as $n \to \infty$ for all $\theta_0 \in \Theta_U$ by Lemma 1. Thus $n^{-1} \sum_{t=1}^{n} (\partial l_t(\theta_0)/\partial\gamma)^2 \xrightarrow{P} 0$ for all $\theta_0 \in \Theta_U$. We also have

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta_i} \frac{\partial l_t(\theta_0)}{\partial \theta_j} &= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{(\nu_0 + 1)b_{0t} - 1}{2} \right)^2 \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \\
&\quad + \frac{1}{n} \sum_{t=1}^{n} \left(\frac{(\nu_0 + 1)b_{0t} - 1}{2} \right)^2 u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \\
&\xrightarrow{P} \frac{(\nu_0 + 1)^2}{4} \mathrm{Var}(b_{0t}) \mathbb{E}[u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0)] = 0
\end{aligned}$$

for all $\theta_0 \in \Theta_U$, $i = 4$, and $j = 2, 3$ by Lemma 7 (i)(ii)(vii). Analogous derivations show that

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \delta} \frac{\partial l_t(\theta_0)}{\partial \nu} \xrightarrow{P} 0$$

for all $\theta_0 \in \Theta_U$. Next, for $i = 2, 3, 4$, we have

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta_i} \frac{\partial l_t(\theta_0)}{\partial \gamma} &= \frac{1}{4n} \sum_{t=1}^{n} ((\nu_0 + 1)b_{0t} - 1)^2 \left(h_{\theta_i t}(\theta_0) h_{\gamma t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\gamma t}(\theta_0) \right) \\
&\quad + \frac{\nu_0 + 1}{2n} \sum_{t=1}^{n} ((\nu_0 + 1)b_{0t} - 1) \, h_{\theta_i t}(\theta_0) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}}.
\end{aligned}$$

By (D.11) of Lemma 7 (vii), the first term converges in L^1 to zero for all $\theta_0 \in \Theta_U$. The second term also converges in L^1 to zero for all $\theta_0 \in \Theta_U$ by the Minkowski and Hölder inequalities, Lemma 2 (i), and Lemma 1. Thus $n^{-1} \sum_{t=1}^{n} (\partial l_t(\theta_0)/\partial\theta_i)(\partial l_t(\theta_0)/\partial\gamma) \xrightarrow{P} 0$ for all $\theta_0 \in \Theta_U$ and $i = 2, 3, 4$. Finally, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial \nu}\frac{\partial l_t(\theta_0)}{\partial \gamma} \\
&\quad= \frac{1}{4n}\sum_{t=1}^{n}\left\{\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\right. \\
&\qquad\qquad \left.\times\left((\nu_0+1)b_{0t}-1\right)\left(h_{\gamma t}(\theta_0)-u_{\gamma t}(\theta_0)\right)\right\} \\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\gamma t}(\theta_0)h_{\nu t}(\theta_0)-u_{\gamma t}(\theta_0)u_{\nu t}(\theta_0)\right) \\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left\{\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\right. \\
&\qquad\qquad\qquad \left.\times\left(\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}\right)\right\} \\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\nu t}(\theta_0)-u_{\nu t}(\theta_0)\right)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}} \\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\times \\
&\qquad\quad \times\left((\nu_0+1)b_{0t}-1\right)u_{\gamma t}(\theta_0) \\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2 u_{\gamma t}(\theta_0)u_{\nu t}(\theta_0) \\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)u_{\nu t}(\theta_0)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}.
\end{aligned}
$$

Note that $u_{\gamma t}(\theta_0)=0$ for all $t\in\mathbb{N}$ and $\theta_0\in\Theta$ by definition. By (D.9) and (D.11) of Lemma 7 (vii), the first, second, and fourth terms converge in L^1 to zero. The third and seventh terms converge in L^1 to zero by the Minkowski and Hölder inequalities, Lemma 1, and Lemma 7 (ii). Thus we have shown that $n^{-1}\sum_{t=1}^{n}(\partial l_t(\theta_0)/\partial\nu)(\partial l_t(\theta_0)/\partial\gamma)\xrightarrow{P}0$ for all $\theta_0\in\Theta_U$.

■

Next, we aim to show that the elements of $\nabla^3_\theta L_n(\theta)$ and $\nabla^3_{\theta^*}L_n(\theta^*)$ are bounded by some stationary processes for all $n\in\mathbb{N}_{>0}$ and $\theta_0,\theta\in\Theta$. For this purpose, we introduce the stationary process $w_t(\theta_0)$ defined by

$$\begin{aligned} w_t(\theta_0) = 1 &+ \frac{\delta_u - \delta_0}{\delta_0}\frac{1}{1-\beta_u} + \frac{\beta_u - \beta_0}{\beta_l}\sum_{k=1}^{t}\prod_{j=1}^{k}\frac{\beta_u}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \\ &+ \frac{\alpha_u(\nu_u+1)}{\beta_l(\nu_l-2)}\sum_{k=1}^{t}\frac{1}{b_{0t-k}}\left(|z_{t-k}| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2\prod_{j=1}^{k}\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t-k}}, \end{aligned} \tag{D.26}$$

for any $0 < \beta_l \leq \beta \leq \beta_u < 1$ and $\theta_0 \in \Theta$, where $\bar{g} \equiv \max\{\gamma_u - \gamma_0, \gamma_0 - \gamma_l\}$.

LEMMA 12. *There exists $\beta_u \in (0,1)$ such that $w_t(\theta_0)$ is strictly stationary and ergodic for each $p \geq 1$, $t \in \mathbb{N}$, $0 < \beta_l \leq \beta \leq \beta_u < 1$, and $\theta_0 \in \Theta$.*

Proof. First, note that

$$\begin{aligned} &\mathbb{E}\left[\max\left\{0, \log\left(\frac{1}{b_{0t}}\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2\right)\right\}\right] \\ &\leq \mathbb{E}\left[\max\left\{0, -\log\left(b_{0t}\right)\right\}\right] + \mathbb{E}\left[\max\left\{0, 2\log\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)\right\}\right] \\ &< \infty \end{aligned}$$

for all $t \in \mathbb{N}$ by the property of the beta and Student's t distributions. Moreover,

$$\mathbb{E}\left[\log\left(\frac{\beta}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}}\right)\right] \leq \mathbb{E}\left[\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t}}\right] < 1$$

for some $\beta_u \in (0,1)$ as b_{0t} is non-degenerate for all $t \in \mathbb{N}$. Thus, the proof is complete by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).

■

In order to show that the elements of $\nabla^3_{\theta} L_n(\theta)$ and $\nabla^3_{\theta^*} L_n(\theta^*)$ are bounded by some stationary processes for all $n \in \mathbb{N}_{>0}$ and $\theta_0, \theta \in \Theta$, we show in Lemma 15 that $h_{\theta_i}(\theta)$, $h_{\theta_i\theta_j}(\theta)$, and $h_{\theta_i\theta_j\theta_k}(\theta)$ are bounded by some stationary processes for all t, $\theta, \theta_0 \in \Theta$, and $i, j, k = 1, \ldots, 5$. In order to show Lemma 15, we use the properties of $h_t(\theta)$ and h_{0t} shown in Lemmas 13 and 14. Lemma 14 is the only place where we use the unit upper-bound on β.

LEMMA 13. *For all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have*

$$\begin{aligned} h_t(\theta) = h_{0t} &+ (\delta - \delta_0)\sum_{k=0}^{t-1}\beta^k + (\beta - \beta_0)\sum_{k=1}^{t}\beta^{k-1}h_{0t-k} \\ &+ \sum_{k=1}^{t}\beta^k\Big[\alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta) - \alpha_0(\nu_0+1)b_{0t-k}h_{0t-k}\Big], \end{aligned} \tag{D.27}$$

$$
\begin{aligned}
h_{0t} = h_t(\theta) + (\delta_0 - \delta) \sum_{k=0}^{t-1} \beta_0^k + (\beta_0 - \beta) \sum_{k=1}^{t} \beta_0^{k-1} h_{t-k}(\theta) \\
+ \sum_{k=1}^{t} \beta_0^k \Big[\alpha_0(\nu_0+1) b_{0t-k} h_{0t-k} - \alpha(\nu+1) b_{t-k}(\theta) h_{t-k}(\theta) \Big].
\end{aligned}
\tag{D.28}
$$

Proof. Since $\delta_0 = h_{0t} - \beta_0 h_{0t-1} - \alpha_0(\nu_0+1) b_{0t-1} h_{0t-1}$, adding and subtracting δ_0 in the equation for $h_t(\theta)$ give

$$
\begin{aligned}
h_t(\theta) = \delta - \delta_0 + \beta h_{t-1}(\theta) + h_{0t} - \beta_0 h_{0t-1} - \alpha_0(\nu_0+1) b_{0t-1} h_{0t-1} \\
+ \alpha(\nu+1) b_{t-1}(\theta) h_{t-1}(\theta).
\end{aligned}
$$

Then (D.27) follows by noting that $h_0(\theta) = h_{00} = \omega_0$. Similarly,

$$
\begin{aligned}
h_{0t} = \delta_0 - \delta + \beta_0 h_{0t-1} + h_t(\theta) - \beta h_{t-1}(\theta) - \alpha(\nu+1) b_{t-1}(\theta) h_{t-1}(\theta) \\
+ \alpha_0(\nu_0+1) b_{0t-1} h_{0t-1}.
\end{aligned}
$$

Then (D.28) follows. ■

LEMMA 14. *For all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have*

(i)

$$
0 \leq \frac{b_t(\theta) h_t(\theta)}{h_{0t}} \leq \frac{1}{\nu - 2} \left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}} \right)^2.
$$

(ii) Define $q_t \equiv q_t(\theta_0) \equiv 1\!\!1_{\{z_t \geq g_u\}}(z_t + g_u/\sqrt{\delta_0})^2 + 1\!\!1_{\{z_t \leq g_l\}}(z_t + g_l/\sqrt{\delta_0})^2$, where $g_u \equiv \gamma_0 - \gamma_l$ and $g_l \equiv \gamma_0 - \gamma_u$. Then

$$
0 \leq \frac{q_t(\theta_0)}{q_t(\theta_0) + (\nu_u - 2) h_t(\theta)/h_{0t}} \leq b_t(\theta) \leq \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2) h_t(\theta)/h_{0t}} \leq 1
$$

a.s. for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$.

(iii)

$$
0 < h_t(\theta)/h_{0t} \leq w_t(\theta_0) \tag{D.29}
$$

a.s. for some strictly stationary process $w_t(\theta_0)$ for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Moreover,

$$
0 < x_t(\theta_0) \leq h_t(\theta)/h_{0t} \tag{D.30}
$$

a.s. for some strictly stationary process $x_t(\theta_0)$ for all $t \in \mathbb{N}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

(iv) $0 \leq \underline{b}_t(\theta_0) \leq b_t(\theta)$, where $\underline{b}_t(\theta_0)$ is some strictly stationary process, for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Moreover, $0 \leq b_t(\theta) \leq \bar{b}_t(\theta_0) \leq 1$, where $\bar{b}_t(\theta_0)$ is some strictly stationary processes, for all $t \in \mathbb{N}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

Proof. (i)-(ii) Note that we have

$$q_t \leq \left(z_t + \frac{g}{\sqrt{h_{0t}}}\right)^2 \leq \left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2. \tag{D.31}$$

Then, for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have a.s.

$$\begin{aligned}\frac{b_t(\theta)h_t(\theta)}{h_{0t}} &= \frac{\left(z_t + g/\sqrt{h_{0t}}\right)^2 h_t(\theta)/h_{0t}}{\left(z_t + g/\sqrt{h_{0t}}\right)^2 + (\nu-2)h_t(\theta)/h_{0t}} \\ &\leq \frac{1}{\nu-2}\left(z_t + \frac{g}{\sqrt{h_{0t}}}\right)^2 \leq \frac{1}{\nu_l-2}\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2.\end{aligned}$$

Likewise, (ii) also follows from (D.31).

(iii) First, we show (D.29). By Lemma 13, we obtain, for all $\theta, \theta_0 \in \Theta$ and $t \in \mathbb{N}$,

$$\begin{aligned}\frac{h_t(\theta)}{h_{0t}} &= 1 + \frac{\delta-\delta_0}{h_{0t}}\sum_{k=0}^{t-1}\beta^k + \frac{\beta-\beta_0}{\beta}\sum_{k=1}^{t}\prod_{j=1}^{k}\frac{\beta h_{0t-j}}{h_{0t-j+1}} \\ &\quad + \frac{\alpha_0(\nu_0+1)}{\beta}\sum_{k=1}^{t}\beta^k\frac{b_{0t-k}h_{0t-k}}{h_{0t}}\left[\frac{\alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta)}{\alpha_0(\nu_0+1)b_{0t-k}h_{0t-k}} - 1\right] \\ &\leq 1 + \frac{\delta_u-\delta_0}{\delta_0}\frac{1}{1-\beta_u} + \frac{\beta_u-\beta_0}{\beta_l}\sum_{k=1}^{t}\prod_{j=1}^{k}\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t-j}} \\ &\quad + \frac{\alpha_u(\nu_u+1)}{\beta_l(\nu_l-2)}\sum_{k=1}^{t}\frac{1}{b_{0t-k}}\left(|z_{t-k}| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2\prod_{j=1}^{k}\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t-j}} \\ &= w_t(\theta_0),\end{aligned}$$

where the inequality in the middle is by Lemma 14 (i).

Next, we show (D.30). By Lemma 14 (ii) and the just derived inequality, we have

$$b_t(\theta) \geq \frac{q_t(\theta_0)}{q_t(\theta_0) + (\nu_u-2)w_t(\theta_0)} \equiv \underline{b}_t(\theta_0) \geq 0,$$

where the process $\underline{b}_t(\theta_0)$ is in terms of the i.i.d. process z_t for any $\theta_0 \in \Theta$ and $t \in \mathbb{N}$. $\underline{b}_t(\theta_0)$ is strictly stationary and ergodic by Lemma 12 and Theorem 3.5.8 of Stout (1974) [also see the relevant results in Royden (1988, p.66-68)]. Note that, by (D.28) of Lemma 13,

$$\begin{aligned}\frac{h_{0t}}{h_t(\theta)} &= 1 + \frac{\delta_0-\delta}{h_t(\theta)}\sum_{k=0}^{t-1}\beta_0^k + \frac{\beta_0-\beta}{\beta_0}\sum_{k=1}^{t}\beta_0^k\frac{h_{t-k}(\theta)}{h_t(\theta)} \\ &\quad + \sum_{k=1}^{t}\frac{\beta_0^k}{h_t(\theta)}\left[\alpha_0(\nu_0+1)b_{0t-k}h_{0t-k} - \alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta)\right].\end{aligned}$$

Since $a_{t-k}/a_t = \prod_{j=1}^k a_{t-j}/a_{t-j+1}$ for any sequence $(a_t)_{t\in\mathbb{N}}$ and $0 < k < t$, we get

$$\begin{aligned} 0 &\leq \frac{h_{0t}}{h_t(\theta)} \\ &= \left|1 + \frac{\delta_0 - \delta}{h_t(\theta)} \sum_{k=0}^{t-1} \beta_0^k + \frac{\beta_0 - \beta}{\beta_0} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0 h_{t-j}(\theta)}{h_{t-j+1}(\theta)} - \alpha(\nu+1) \sum_{k=1}^{t} b_{t-k}(\theta) \prod_{j=1}^{k} \frac{\beta_0 h_{t-j}(\theta)}{h_{t-j+1}(\theta)}\right| \\ &\div \left|1 - \alpha_0(\nu_0+1) \sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0 h_{0t-j}}{h_{0t-j+1}}\right| \end{aligned} \tag{D.32}$$

The numerator of (D.32) is bounded above by

$$\begin{aligned} 1 &+ \frac{\delta_u - \delta_l}{\delta_l(1-\beta_u)} + \frac{\beta_u - \beta_l}{\beta_0} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)} \\ &+ \alpha_u(\nu_u+1) \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)}. \end{aligned}$$

Since $\underline{b}_t(\theta_0)$ is non-degenerate, strictly stationary and ergodic, there exists $\beta_l \in (0,1)$ such that this quantity is strictly stationary and ergodic by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). In the denominator of (D.32), we have

$$\sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0 h_{0t-j}}{h_{0t-j+1}} = \sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0}{\delta_0/h_{0t-j} + \beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \tag{D.33}$$

Since $\max\{0, \log|X|\} \leq |X|$ for any real valued random variable X, we have $\mathbb{E}[\max\{0, |b_{0t}|\}] \leq \mathbb{E}[|b_{0t}|] < 1$ for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Moreover,

$$\mathbb{E}\left[\log\left|\frac{\beta_0}{\delta_0/h_{0t} + \beta_0 + \alpha_0(\nu_0+1)b_{0t}}\right|\right] < 0$$

for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Thus, the denominator of (D.32) is strictly stationary and ergodic if $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).[5] Thus, we have found a strictly stationary process $x_t(\theta_0)$ such that $0 < x_t(\theta_0) \leq h_t(\theta)/h_{0t}$ for all $t \in \mathbb{N}$, $\theta \in \Theta$ and $\theta_0 \in \Theta_L$.

(iv) By Lemma 14 (iii), we obtained $b_t(\theta) \geq \underline{b}_t(\theta_0) \geq 0$. Moreover, by Lemma

[5] This is the only place that limits us from proving the consistency and asymptotic normality results for the nonstationary case (i.e. when $\theta_0 \in \Theta_U$). In order to show the asymptotic properties for $\theta_0 \in \Theta_U$, we would need to find a strictly stationary and ergodic process that bounds the denominator of (D.32) from below. We find showing this difficult when the RHS of (D.33) is greater than or equal to one.

14 (ii)(iii), we have

$$b_t(\theta) \leq \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2)h_t(\theta)/h_{0t}}$$
$$\leq \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2)x_t(\theta_0)} \equiv \bar{b}_t(\theta_0),$$

where, for all $\theta \in \Theta$ and $\theta_0 \in \Theta_L$, $\bar{b}_t(\theta) \in [0, 1]$ and $(\bar{b}_t(\theta))_{t\in\mathbb{N}}$ is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974) and Royden (1988, p.66-68). ■

Finally, in order to show Lemma 15, define the following process.

$$u^*_{\theta_i t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i}}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i = 1, \ldots, 5$, where

$$m_t(\theta_0) \equiv \max\left\{\frac{\beta_u + \alpha_u(\nu_u + 1)\underline{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l + 1)\underline{b}_t(\theta_0)}, \frac{\beta_u + \alpha_u(\nu_u + 1)\bar{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l + 1)\bar{b}_t(\theta_0)}\right\}$$

is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974) and Royden (1988, p.66-68). $\widetilde{u}_{\theta_i}$ bounds $\widehat{h}_{\theta_i t}(\theta)$ for all t, $\theta \in \Theta$, and $i = 1, \ldots, 5$. We set

$$\widetilde{u}_\delta = 1/\delta_l, \qquad \widetilde{u}_\alpha = \nu_u + 1, \qquad \widetilde{u}_\beta = 1,$$
$$\widetilde{u}_\gamma = 2\alpha_u(\nu_u + 1)\max\{1, ((\nu_l - 2)\delta_l)^{-1}\},$$
$$\widetilde{u}_\nu = \alpha_u + \alpha_u(\nu_u + 1)(\nu_l - 2)^{-1}.$$

Moreover, define the following process;

$$u^*_{\theta_i\theta_j t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i, j = 1, \ldots, 5$, where $\widetilde{u}_{\theta_i\theta_j t}(\theta_0)$ bounds $\widehat{h}_{\theta_i\theta_j t}(\theta)$ for any t, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$. They are defined in Appendix B.2. Similarly, we define

$$u^*_{\theta_i\theta_j\theta_m t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i\theta_j\theta_m t-k}(\theta_0)}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i, j, m = 1, \ldots, 5$, where $\widetilde{u}_{\theta_i\theta_j\theta_m t}(\theta_0)$ bounds $\widehat{h}_{\theta_i\theta_j\theta_m t}(\theta)$ for any t, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$. For instance, for $i = j = m = 3$ (i.e. $\theta_i = \theta_j = \theta_m = \beta$), we set

$$\widetilde{u}_{\beta\beta\beta t}(\theta_0) = 3u^*_{\beta\beta t}(\theta_0) + 6\alpha_u(\nu_u + 1)u^*_{\beta t}(\theta_0)^3 + 2\alpha_u(\nu_u + 1)u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0).$$

Lemma 15 establishes some of the useful properties of these processes.

LEMMA 15. *For all $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i, j, m = 1, \ldots, 5$;*

(i) $|h_{\theta_i t}(\theta)| \leq u^*_{\theta_i t}(\theta_0)$ for all $t \in \mathbb{N}$.

(ii) $|h_{\theta_i\theta_j t}(\theta)| \leq u^*_{\theta_i\theta_j t}(\theta_0)$ *for all* $t \in \mathbb{N}$.

(iii) $|h_{\theta_i\theta_j\theta_m t}(\theta)| \leq u^*_{\theta_i\theta_j\theta_m t}(\theta_0)$ *for all* $t \in \mathbb{N}$.

(iv) $\left(u^*_{\theta_i t}(\theta_0)\right)_{t\in\mathbb{N}}$, $\left(u^*_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$, *and* $\left(u^*_{\theta_i\theta_j\theta_m t}(\theta_0)\right)_{t\in\mathbb{N}}$ *are strictly stationary and ergodic.*

Proof.

(i) It is easy to show that $\left|\widehat{h}_{\theta_i t}(\theta)\right| < \widetilde{u}_{\theta_i}$ for all t and $i = 1, \ldots, 5$. Note that $|\widehat{h}_{\gamma t}(\theta)| < \widetilde{u}_{\gamma}$ can be verified by the condition, (D.1), of Lemma 1. Then, by the condition, (D.3), of Lemma 2 (i), we obtain

$$
\begin{aligned}
|h_{\theta_i t}(\theta)| &\leq \sum_{k=1}^{t} \frac{|\widehat{h}_{\theta_i t}(\theta)|}{\beta_l} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta + \alpha(\nu+1)b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)} \\
&\leq \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i}}{\beta_l} \prod_{j=1}^{k} \frac{\beta_u + \alpha_u(\nu_u+1)b_{t-j}(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)b_{t-j}(\theta_0)} \\
&\leq u^*_{\theta_i t}(\theta_0),
\end{aligned}
$$

where the last inequality used the fact that

$$
\begin{aligned}
\frac{\beta + \alpha(\nu+1)b_t(\theta)^2}{\beta + \alpha(\nu+1)b_t(\theta)} &\leq \max\left\{ \frac{\beta_u + \alpha_u(\nu_u+1)\underline{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)}, \frac{\beta_u + \alpha_u(\nu_u+1)\bar{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\bar{b}_t(\theta_0)} \right\} \\
&\equiv m_t(\theta)
\end{aligned}
$$

for all t, $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i = 1, \ldots, 5$ by Lemma 14 (iv).

(ii) Derivations analogous to Lemma 15 (i) show that $|h_{\theta_i\theta_j t}(\theta)| \leq u^*_{\theta_i\theta_j t}(\theta_0)$ for all t, $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i, j = 1, \ldots, 5$. Note that we can verify $|\widehat{h}_{\theta_i\theta_j t}(\theta)| < \widetilde{u}_{\theta_i\theta_j t}(\theta_0)$ whenever $i = 4$ or $j = 4$ (i.e. $\theta_i = \theta_4 = \gamma$ or $\theta_j = \gamma$) by the condition, (D.1), of Lemma 1 and by noting that

$$
\left|\frac{b_t(\theta)}{e_t}\right| = \frac{|e_t|}{e_t^2 + (\nu-2)h_t(\theta)},
$$

which is bounded a.s. by Lemma 1.

(iii) This proof is analogous to the proofs for Lemma 15 (i)(ii).

(iv) $(m_t(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary, and we can find $(\beta_u, \beta_l, \alpha_u, \alpha_l, \nu_u, \nu_l)$ such that $\mathbb{E}\left[\ln\left(m_t(\theta_0)\right)\right] < 0$ for all $\theta_0 \in \Theta_L$ since $\bar{b}(\theta_0) \in (0,1)$ and $\underline{b}(\theta_0) \in (0,1)$ are non-degenerate. Moreover, using the property that $\ln(x) \leq x - 1$ for all $x > 0$, we have

$$
\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i}|\}\right] \leq \mathbb{E}\left[\max\{0, |\widetilde{u}_{\theta_i}|\}\right] = \mathbb{E}\left[|\widetilde{u}_{\theta_i}|\right] < \infty
$$

for $i = 1, \ldots, 5$. Then $(u_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).

Likewise, we can show that $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j t}(\theta_0)|\}\right] < \infty$ and $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j\theta_m t}(\theta_0)|\}\right] < \infty$ for all t, $\theta_0 \in \Theta$, and $i, j, m = 1, \ldots, 5$. Then we can deduce that $(u^*_{\theta_i\theta_j t}(\theta_0))_{t\in\mathbb{N}}$ and $(u^*_{\theta_i\theta_j\theta_m t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for any $\theta_0 \in \Theta_L$ and $i, j, m = 1, \ldots, 5$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). ■

We are now ready to show that the elements of $\nabla^3_\theta L_n(\theta)$ are bounded by some stationary and ergodic sequence for all $n \in \mathbb{N}_{>0}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

LEMMA 16. *For any $\theta_0 \in \Theta_L$ and $n \in \mathbb{N}$, we have*

$$\max_{i,j,m=1,\ldots,5} \sup_{\theta\in\Theta} \left|\partial^3 L_n(\theta)/\partial\theta_i\partial\theta_j\partial\theta_m\right| \leq c_n,$$

where $0 \leq c_n \xrightarrow{P} c$ as $n \to \infty$ and $0 < c < \infty$.

Proof. For the third derivative with respect to β, from (B.2), we have

$$\begin{aligned}
\left|\partial^3 L_n(\theta)/\partial\beta^3\right| &= \frac{1}{n}\left|\sum_{t=1}^n \frac{\partial^3 l_t(\theta)}{\partial\beta^3}\right| \\
&\leq \frac{1}{n}\sum_{t=1}^n (\nu+1) b_t(\theta)(1-b_t(\theta)) \left[\left|h_{\beta t}(\theta)^3\right| + \frac{3}{2}\left|h_{\beta t}(\theta) h_{\beta\beta t}(\theta)\right|\right] \\
&\quad + \frac{1}{n}\sum_{t=1}^n (\nu+1)\left|h_{\beta t}(\theta)^3\right| b_t(\theta)(1-b_t(\theta))^2 \\
&\quad + \frac{1}{2n}\sum_{t=1}^n \left(3\left|h_{\beta t}(\theta)h_{\beta\beta t}(\theta)\right| + 2\left|h_{\beta t}(\theta)^3\right| + \left|h_{\beta\beta\beta t}(\theta)\right|\right)\left[(\nu+1)b_t(\theta) - 1\right] \\
&\leq \frac{\nu_u+1}{n}\sum_{t=1}^n \left(2u^*_{\beta t}(\theta_0)^3 + \frac{3}{2}u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0)\right) \\
&\quad + \frac{\nu_u+2}{2n}\sum_{t=1}^n \left(3u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0) + 2u^*_{\beta t}(\theta_0)^3 + u^*_{\beta\beta\beta t}(\theta_0)\right) \\
\xrightarrow{P}\ & (\nu_u+1)\left(2\mathbb{E}\left[u^*_{\beta 1}(\theta_0)^3\right] + \frac{3}{2}\mathbb{E}\left[u^*_{\beta 1}(\theta_0)u^*_{\beta\beta 1}(\theta_0)\right]\right) \\
&+ \frac{\nu_u+2}{2}\left(3\mathbb{E}\left[u^*_{\beta 1}(\theta_0)u^*_{\beta\beta 1}(\theta_0)\right] + 2\mathbb{E}\left[u^*_{\beta 1}(\theta_0)^3\right] + \mathbb{E}\left[u^*_{\beta\beta\beta 1}(\theta_0)\right]\right) \\
\in\ & (0, \infty)
\end{aligned}$$

as $n \to \infty$ by Lemma 15(i)-(iv) and Theorem 3.5.7 of Stout (1974). Straightforward differentiation shows that the desired inequality holds for other third derivatives by Lemma 15(i)-(iv) and Lemma 1. ■